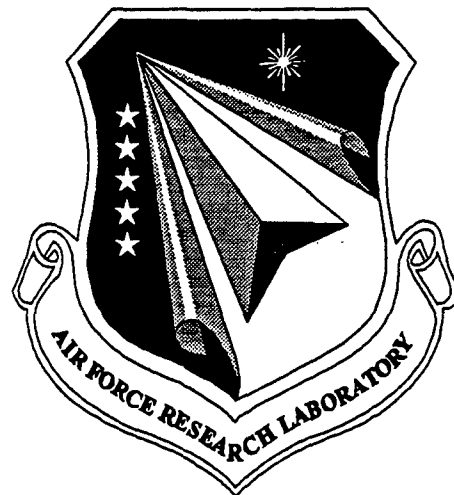AFRL-ML-WP-TR-1999-4033

INTEGRATED SUBSTRATE AND THIN FILM DESIGN METHODS

STEPHEN L. THALER

IMAGINATION ENGINES, INC.
12906 AUTUMN VIEW DRIVE
ST. LOUIS, MO 63146-4331

FEBRUARY 1999

19991104 014

FINAL REPORT FOR 05/01/1998 – 02/01/1999

THIS IS A SMALL BUSINESS INNOVATION RESEARCH (SBIR) PHASE I REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

MATERIALS AND MANUFACTURING DIRECTORATE
AIR FORCE RESEARCH LABORATORY
AIR FORCE MATERIEL COMMAND
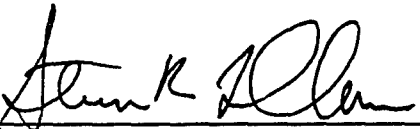WRIGHT-PATTERSON AIR FORCE BASE OH 45433-7750

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto. ·

This report has been reviewed by the Office of Public Affairs (ASC/PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

CAPT DAVID CONRAD, USAF
Project Engineer
Materials Process Design Branch
Manufacturing Technology Division

DR. STEVEN R. LeCLAIR
Chief
Materials Process Design Branch
Manufacturing Technology Division

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	FEBRUARY 1999	FINAL REPORT FOR 05/01/1998 - 02/01/1999

4. TITLE AND SUBTITLE
INTEGRATED SUBSTRATE AND THIN FILM DESIGN METHODS

5. FUNDING NUMBERS
C   F33615-98-C-5139
PE 65502
PR 3005
TA 05
WU AR

6. AUTHOR(S)

STEPHEN L. THALER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
IMAGINATION ENGINES, INC.
12906 AUTUMN VIEW DRIVE
ST. LOUIS, MO 63146-4331

8. PERFORMING ORGANIZATION REPORT NUMBER

001AE

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
MATERIALS AND MANUFACTURING DIRECTORATE
AIR FORCE RESEARCH LABORATORY
AIR FORCE MATERIEL COMMAND
WRIGHT-PATTERSON AFB, OH 45433-7750
POC: CAPT DAVID CONRAD , USAF, AFRL/MLMR, 937-255-8786

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AFRL-ML-WP-TR-1999-4033

11. SUPPLEMENTARY NOTES

THIS IS A SMALL BUSINESS INNOVATION RESEARCH (SBIR) PHASE I REPORT

12a. DISTRIBUTION AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE, DISTRIBUTION UNLIMITED

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report was developed under SBIR contract.

An artificial neural network cascade, containing 16 individual network modules and approximately 1,000 processing units, has produced an interactive database of nearly a quarter million potential binary and ternary chemical systems. While many of these hypothetical materials are anticipated to be thermodynamically stable, they are most likely kinetically inaccessible via typical bulk chemistry routes. However, since modern thin film technology allows a wide range of exotic compositions and stoichiometries via deposition, surface treatments, and nano-fabrication, it is anticipated that this newly acquired theoretical database will form a comprehensive roadmap to the formation of previously unattainable materials that offer significant technological advantages. Further, with the suite of available coating materials greatly expanded, thin film designers now have at their disposal the means to implement multilayer and composite thin film device designs that fulfill a much broader range of performance requirements and that are ideally matched to both underlying substrate and external working environment.

14. SUBJECT TERMS
SBIR Report, Autonomous Discover, Neural Networks, Thin Films, Materials, Binary Chemical Systems, Ternary Chemical Systems

15. NUMBER OF PAGES
53

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

TABLE OF CONTENTS

Integrated Substrate and Thin Film Design Methods

Abstract – An artificial neural network cascade, containing 16 individual neural network modules and approximately 1,000 processing units, has produced an interactive database containing a broad range of chemical and physical properties for nearly a quarter million potential binary and ternary chemical systems. While many of these hypothetical materials are anticipated to be thermodynamically stable, they are most likely kinetically inaccessible via typical bulk chemistry routes. However, since modern thin film technology allows a wide range of exotic compositions and stoichiometries via deposition, surface treatments, and nano-fabrication, it is anticipated that this newly acquired theoretical database will form a comprehensive roadmap to the formation of previously unattainable materials that offer significant technological advantages. Further, with the suite of available coating materials greatly expanded, thin film designers now have at their disposal the means to implement multi-layer and composite thin film device designs that fulfill a much broader range of performance requirements and that are ideally matched to both underlying substrate and external working environment.

1. Introduction

1.1 Motivation for a Theoretical Database of Filming Materials

The first stage in the design of any thin film device is the identification of the necessary chemical and physical attributes of the constituent films. For instance, in the design of a multi-layer optical stack, one generally begins with a prescription for the optical thickness of each film layer (nd), along with the refractive indices of both substrate and the external medium. Typically, the thin film designer seeks materials that not only satisfy these optical requirements, but likewise offer other favorable characteristics such as chemical stability and wear resistance. From these specifications the designer then chooses from a relatively narrow suite of known optical filming materials. Unfortunately, the lack of filming materials satisfying the overall device formulation may force the total redesign of the thin film structure. Effectively, materials availability currently drives thin film design.

With the development of advanced film growth procedures such as pulsed laser deposition, molecular beam epitaxy, and nano-fabrication techniques, it is now possible to specifically tailor materials to meet general thin film design requirements. The new problem becomes one of theoretically predicting and producing the needed materials, many of which have never been purposely synthesized. Although the condensed matter theoretician may have many predictive tools at his disposal, such as quantum mechanically-based band calculations, such methodologies require painstaking care in set up and significant CPU time to implement. Disappointingly, the resulting calculations apply only to very narrow regimes of composition and phase. What is needed is a sweeping prediction of materials properties without recourse to highly specialized ab initio techniques.

Artificial neural networks afford a convenient means of estimating materials properties. In contrast to traditional computational approaches, neural networks can rapidly form their own models, interrelating attributes of composition and phase to both physical and chemical characteristics. Such neural networks could, for instance, be used to scan large databases of known materials producing estimates of key properties for each formula and phase. The problem with such a data mining approach would be the evaluation of potential yet unsynthesized or unreported chemical compounds and systems that would obviously be absent from such a relatively static database of known materials.

What is needed is a generative technique that may recommend multitudes of potential chemical systems whose materials properties may be subsequently evaluated via readily attainable neural network models. One candidate approach would involve the building of an algorithmic production system that embodies a knowledge base of rules of chemical bonding, valence, charge neutrality, etc. The problem with such a scheme is that the underlying chemical heuristics remain piecemeal at this time. Typically, crystallographic systems are usually rationalized only after their actual discovery, with industrious x-ray crystallographers deducing underlying stoichiometry and structure only after the material is in hand. Occasionally, small substitutional variations are foreseen, but rarely does condensed matter physics theorize and then fabricate such novel crystal structures from scratch.

Alternately, one may consider genetic algorithms (GA) as the route to building such a generative engine. The pitfalls in this methodology, especially when applied to chemical formulation problems, are the risks of combinatorial explosion and the production of genetic mutations that do not necessarily embody chemical constraint relations. Specifically, as genes or alleles randomly combine, there is no foreknowledge by the underlying engine that the resulting formula or structure will indeed be a plausible one. Typically any critics or objective functions are inundated in selecting the most robust species. Finally, there is no facility by which these algorithms may adapt and hence learn from their own mistakes.

A newer alternative (that may in fact readily learn from its experience) is that of the patented "Creativity Machine Paradigm" (Thaler, 1997, U.S. Patent 5,659,666), in which the search engine consists of either an internally or externally perturbed artificial neural network trained upon large numbers of verified chemical species. Exposure of this net to various forms of synaptic noise tends to drive it into activation states representing new and plausible chemical species. Allowed to run in this 'dreaming' mode, immense databases may be populated with materials that have a high likelihood of being synthesized. In contrast to the GA approach, each new species proposed by the perturbed net obeys cumulatively gleaned constraint relations acquired through its training exposure to known chemical systems. Hence chemical attributes mutate in a cooperative way, wherein all 'genes' (i.e., atoms, stoichiometry, and structure) are mutually sensitive to one another's state.

Further, if each species proposed by such a *dreaming* network is evaluated for its anticipated properties by subsidiary neural network modules, a veritable encyclopedia of

potential chemical compounds may be amassed. Now, thin film designers have at their disposal a continuum of materials possibilities to fulfill a wide range of specific device and / or performance requirements. In all likelihood, it is in the thin film laboratory, rather than in a bulk chemistry community, that these novel materials will be realized. This unique advantage stems from the wide range of thin film procedures at their disposal that may rapidly produce quasi-stable materials in small proportions. These techniques include (1) chemical vapor deposition, (2) pulsed laser deposition, (3) molecular-beam epitaxy, (4) ion-beam implantation, (5) laser and x-ray surface conversion and modification, (6) splat-cooling, and (7) spallation, and (8) nano-fabrication techniques involving Scanning Tunneling Electron Microscopy (STEM). Effectively, this suite of thin film processing technique would allow the rapid prototyping of novel materials discovered by a Creativity Machine.

1.2 The Creativity Machine Paradigm
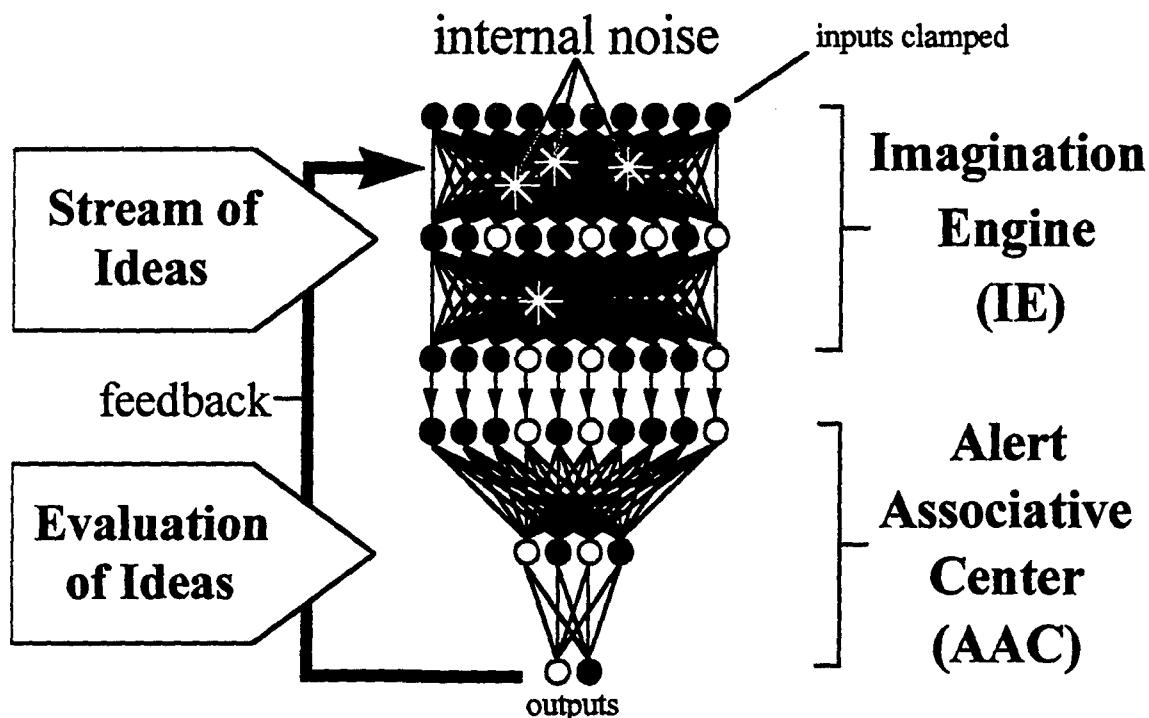
Figure 1. A Simple Creativity Machine. Here, the starbursts represent 'hopping' perturbations among the connection weights of the imagination engine.

Rumelhart and colleagues (1986) emphasized the utility of parallel distributed processing systems as constraint satisfaction networks in their pioneering work. Using "hand-wired" competitive networks exposed to various room schemata, they were able to demonstrate a primitive brand of creativity in which novel, yet plausible furniture combinations were predicted. Using the well-known principle of *vector completion*, the net could accept incomplete inputs representing a partially described room and through subsequent annealing could arrive at a fuller description of that room. Therefore, when only supplied with the inputs of a stove and a coffee cup, the net could finally arrive at a network state

in which additional processing units corresponding to a refrigerator, sink, and oven could be likewise activated. In other words, the net was prescribing a plausible room setting that it may not have "seen" within its training experience. In this sense, such a network was *inventing* new room types.

It has recently been demonstrated (Thaler, 1995, 1996 a, b, c) that a trained artificial neural network supplied no inputs whatsoever, and driven by stochastic perturbations to its internal architecture may generate self-consistent schemata related to the conceptual space embodied within its training exemplars. In short, the network is perceiving something when in fact there are no presented environmental inputs. Accordingly the term "virtual input effect" has been coined to describe the phenomenon. Contacting with Rumelhart's work, if we were to train a simple auto-associative feedforward net on numerous examples of room schemata (hence bypassing the tedious Bayesian statistics used to construct his net), setting the inputs of the network to values of zero and then randomly perturbing the connection weights from their trained values, we would observe a progression of network activations corresponding to plausible furniture schemes. The difference in operating procedure from Rumelhart's work is significant, representing the distinction between *perception* with its processing of environmental features, and *internal imagery* with its inherent independence from such external entities. In the Rumelhart's original work, an associative net is interpreting some partial environmental vector as something it has never seen. In the case of virtual input effect, the net is in a state tantamount to sensory deprivation, in effect hallucinating within a silent and darkened room.

When supplied no external inputs, the production of meaningful activations by the network relies upon a different brand of vector completion than is normally discussed. Rather than fill in incomplete or corrupted input patterns, the net attempts to complete internal, noise-induced activation patterns within the net's encryption layers. Therefore, any local or temporary damage to the network's mapping is interpreted by downstream layers as some "familiar" activation pattern normally encountered upon application of a training exemplar to the network's inputs (Thaler, 1995). Because of the many combinatorial possibilities in perturbing connection weights within a network, we arrive at a means for generating proportionately more novel schema than is possible with input perturbations alone. Furthermore, because the connection traces within a trained neural network generally correspond to the rules binding the underlying conceptual space together, such stochastic perturbation schemes serve to soften these rules, in turn allowing a gradual departure from the known space of possibilities. The result is a strictly neurological search engine whose internal noise level may be parametrically increased to achieve progressively more novel concepts. I call such a chaotic network an *imagination engine* or IE.

By attaching to the IE a critic network (termed an *alert associative center* or AAC) that has been trained by example to recognize any emerging concept that possesses utility or value, a *Creativity Machine* (Figure 1) is formed. Because the only inputs to this closed loop system take the form of unintelligible stochastic perturbations (i.e., heat), the system is deemed autonomous. Therefore, it monitors its own chaotically generated stream of consciousness, if you will, periodically extracting and isolating any concepts offering

usefulness. The critic net may in turn modulate the intensity of perturbation within the first net, willfully dropping the computational temperature within the IE when that network appears to be on the right track (i.e., an attentional mechanism).

Table 1. Some Recent Creativity Machine Successes

Application Area	Outcome	Reference
musical composition	copyrighting of 11,000 novel musical 'hooks'	U.S. Copyright PAu-1-920-845 "Musical Themes From Creativity Machine"
materials discovery	autonomous generation of a materials database, including potentially new ultrahard materials and high-temperature superconductors	Autonomous Materials Discovery via Spreadsheet-Implemented Neural Network Cascades, JOM-e, 49(4) (1997) http://www.tms.org/pubs/journals/JOM/970 4/Thaler
beverage invention	a dynamic database of over 15,000 mixed drinks	http://www.imagination-engines.com/NeuralBar/Nbar.htm
personal hygiene product design	20% improvement in performance over existing designs	Gillette / Oral-B
control system	successful construction and testing of thin film coating reactor control system that invents recovery paths.	U.S. Air Force SBIR contract AF96-152, Automated Data Acquisition For In Situ Material Process Modeling
new supermagnet formulations	significant fabrication cost reduction in FeB-R ceramic supermagnets.	Basic Research Corp.
retail portfolio distribution	Creativity machine successfully recommends shelving distributions as a function of demographics surrounding retail outlet	Anheuser-Busch, Inc.

The practicality and successes (see Table 1 for a few examples) of the Creativity Machine paradigm stem from the fact that all networks involved are trained by example. Therefore, as long as historical data exist within any conceptual space, backpropagation or any other neural network learning paradigm may be used to rapidly train the required Creativity Machine networks. This ease of construction has allowed the building of a wide variety of Creativity Machines focused on diverse knowledge spaces, ranging from music composition, to ultrahard materials discovery, to the invention of personal hygiene products.

Common to the operation of most Creativity Machines built to date is a perturbation scheme in which small disturbances stochastically "hop" among the connection weights of the network. To parametrize the internal chaos within the IE, the governing algorithm parcels out n perturbations, usually of fixed or average magnitude σ, then randomly and cyclically distributes them among the N total connection weights of the IE. In Figure 2, for instance, when network inputs are clamped, the governing algorithm places four perturbations (represented by starbursts) of fixed magnitude at time t_0, resulting in a distinct activation pattern at the network's outputs that represents some idea or concept. On every half cycle, $t_0 + \delta t/2$, the perturbations are removed, restoring the net to its trained-in state. Finally, in initiating a new cycle at time $t_0 + \delta t$, the algorithm randomly places the n perturbations of magnitude σ on newly chosen connection weights. When

viewed as a rapid graphical succession, the hopping motion resembles a boiling liquid, hence suggesting the term "cavitation" to describe this specific agenda of stochastic network perturbation.

$$t_0 \qquad\qquad t_0 + \delta t/2 \qquad\qquad t_0 + \delta t$$
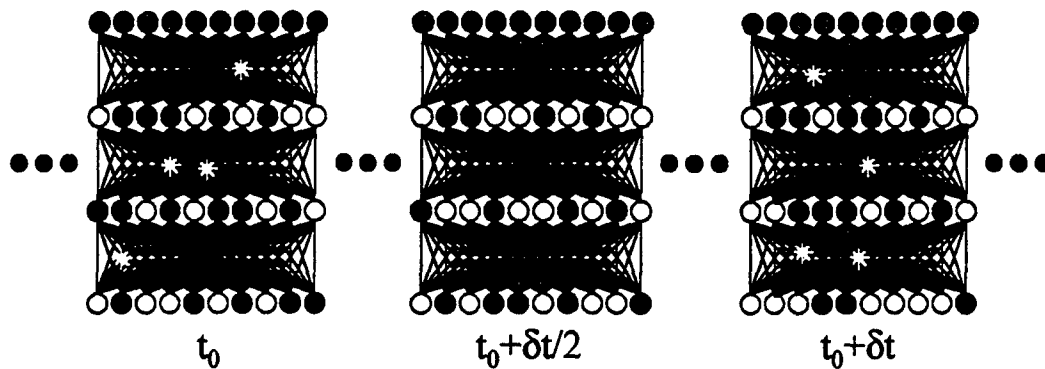
Figure 2. "Cavitation" of the Imagination Engine.

Therefore, during operation the Creativity Machine may be run under a whole range of operating conditions governed by the parameters n, σ, and δt that collectively specify the level of cavitation applied to the IE. Obviously, applying no perturbation at all ($n = 0$ or $\sigma = 0$) to the IE will result in no activation turnover and hence no idea generation. Alternatively, applying large perturbations n and σ will produce such significant degradation to the network mapping that all constraints are destroyed within the captured knowledge domain. The result of severe perturbation is therefore to produce totally unconstrained activation patterns containing little, if any, information content. The former regime consists of vanishingly small perturbations and is regarded as "Neo-Lamarckian" in nature (Rowe and Partridge, 1993), representing a highly constrained and hence inefficient discovery mechanism. The latter unconstrained search regime, at high values of n and σ, is considered "Neo-Darwinian" (i.e., the working regime of genetic algorithms) and is likewise inefficient due to the extensive sifting required by the critic network to find meaningful information among the multitudes of unconstrained concepts produced.

Obviously the ideal regime for Creativity Machine operation lies somewhere between the Neo-Darwinian and Neo-Lamarckian search regimes. To achieve the necessary level of internal perturbation, the parameters n and σ are adjusted so that the quantity $n\sigma/N$ (where N is the total number of connection weights in the IE) is approximately 0.05–0.06, representing the mean perturbation per connection weight in the IE. Dividing through by δt, the perturbation time constant depicted in Figure 2, we obtain a parameter called the "cavitation rate,"

$$\rho = n\sigma/N\delta t , \qquad\qquad (1)$$

representing the mean rate of perturbation for any connection weight in the IE and the single most important controlling parameter of both the quantity and quality of concepts generated by the overall Creativity Machine architecture.

Analytically, the choice of mean perturbation $n\sigma/N$ = 0.06 generally defines a cusp in network behavior that separates a regime of perturbation level corresponding to intact memory recall from that of increasingly corrupted memory generation (i.e., confabulation). This transition in the fidelity of network activations is a generally observed pattern among all IEs used to date and is exemplified in Figure 3, where we see this behavioral transition in a plot of the probability of intact memory production versus *cavitation rate* within a small internally perturbed network with constant inputs. The net has been trained to contain the memory of 16 binary vectors. This distinct separation between intact memory and confabulation persists even within more abstract conceptual spaces that may include subjective areas such as musical composition or more objective problems as in the discovery of new high-temperature superconductors (as discussed in Figure 4).
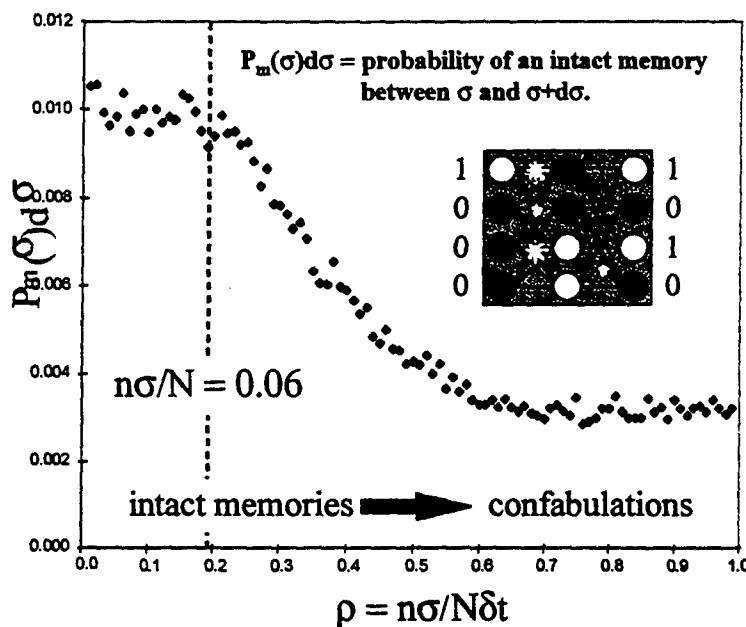
Figure 3. The Probability of the Noise-Induced Activation of an Intact Network Memory as a Function of the Cavitation Rate, ρ. Note the cusp near nσ/N=0.06 dividing intact from corrupted memory recall. The plot is the result of 1,000 cavitation cycles applied to the simple auto-associative net shown in the inset, trained on 16 binary vectors, subjected to n=4 perturbations of variable magnitude σ and a time constant δt of 0.3 sec. Inputs of the net were clamped at the binary memory (1,0,0,0).

We find in general that the most fertile cavitation regime corresponds to mean connection weight perturbations near 0.06. At lower perturbation levels the IE revisits largely training exemplars and their generalizations. At progressively higher levels of connection weight perturbation, the IE produces less constrained and hence more nonsensical possibilities (i.e., noise).

Realizing that the connection weights of a neural network implicitly contain the rules and schema that bind together any given conceptual space, the perturbation scheme embodied within cavitation effectively experiments with these rules by softening them either

individually or in parallel while the AAC judges the utility of the resulting concepts. A mean connection weight perturbation of approximately 0.06 appears to be a universal amount by which to soften these internal rules without producing nonsensical or known concepts. Symbolically representing the constraint relations within any given neural network as the unit sphere, coherent concepts that embody most of the useful ideas emerging from an IE fall within a thin membrane surrounding the $n\sigma/N = 0.06$ surface, no matter what the conceptual space involved. Excursions too far beyond this surface, where $n\sigma/N \gg 0.06$, generally produce nonsense, as intimated in Figure 4.

● = training exemplar
○ = invented
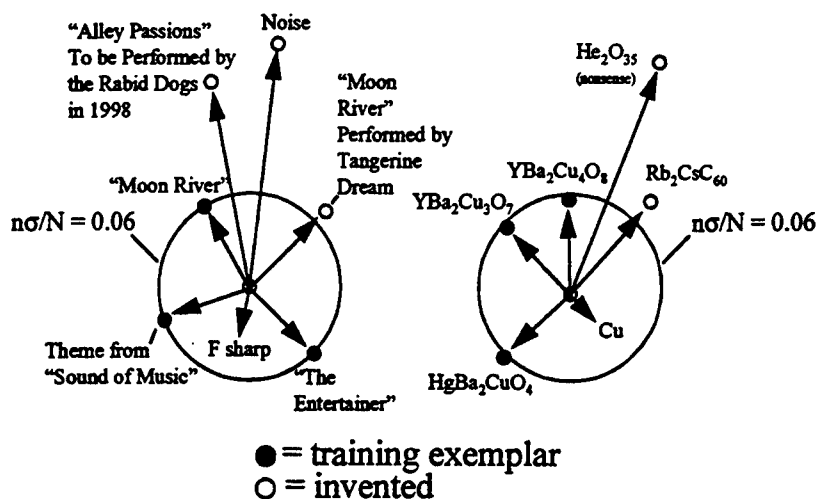
Figure 4. The invention of a plausible concept by the imagination engine takes place within a membrane surrounding the $n\sigma/N = 0.06$ surface, corresponding to the cusp region in Figure 3. Excursions in mean connection weight perturbation significantly beyond this regime produce noise.

1.3 Creativity Machine Families Applied to Materials Discovery

Many variations exist on the patented Creativity Machine (CM) architecture, but typically four main families have been described. These categories are elaborated below, in the specific context of materials discovery applications.

1.3.1 Type I Creativity Machine

This CM class is distinguished by the fact that noise is applied only to input nodes of an IE network to implement a purely stochastic search, based entirely upon network outputs (i.e., no separate critic net is utilized). Such imagination engines are useful in solving problems in which there are no mutual constraints among the perturbed input attributes. Such stochastic nets are used in mixture problems, wherein there are few restrictions upon the proportions of any given ingredient (i.e., paints). Characteristically, the output(s) of such an IE are patrolled for maxima or minima in certain mixture attributes, until the uphill climb results in an optimal materials formulation. Obviously, the Type I CM is ill-suited to stoichiometric chemical problems due to the complex chemical constraints involved in determining fitting stoichiometric indices x, y, and z for any combination of elements A, B, and C.

1.3.2 Type II Creativity Machine

When complex relations exist among the attributes of a given problem, the Type II imagination engine is called into play. Typically the imagination engine consists of an auto-associative network that tends to classify internal synaptic perturbation as either (1) training exemplars (i.e., memory recall) or (2) novel patterns, representing plausible generalizations of the training exemplars. In single-pass operation of such IE nets, feedforward propagation, driven by noise, produces novel possibilities that are then filtered for utility by an attached neural network critic or algorithm. The perturbation scheme used involves the network cavitation process embodied in Equation 1. Typically the phase boundary between memory recall and useful confabulation falls at the mean synaptic perturbation level of $n\sigma/N = 0.06$ Such Type II CMs are likewsie inappropriate to stoichiometric discovery since single pass operation tends to deliver only approximations to potential chemical compositions. That is, a single feedforward operation does not allow the net to seek and completely settle into an appropriate attractor basin representing a potential chemical species.

1.3.3 Type III Creativity Machine

If a Type II auto-associative IE is made recurrent, repetitive feedforward propagation of any internal network perturbation tends to push the network activation state into various attractor basins representing either (1) memories or (2) confabulations that exercise the learned constraints relations among the problem attributes. A network critic or algorithm may monitor the emerging memories and confabulations for utility or value. Such networks are ideal for capturing the myriad constraint relations among constituent elements A, B, and C and their respective stoichiometries x, y , and z. Furthermore, multiple stoichiometries x, y, and z for fixed elements A, B, and C may be accessed by such nets, owing to the many-to-many mapping afforded by auto-associative nets. Therefore, such a net may visit attractor basins representing H_2O_2, as well as H_2O.

Important to realize in the function of a Type III Creativity Machine applied to stoichiometric discovery, is that the imagination engine works as an auto-associative memory that begins operation with pinned inputs representing, for instance, fixed elements A, B, and C. The remaining inputs for the stoichiometries x, y, and z remain free to vary. By applying random numbers (i.e., noise) to these free stoichiometric nodes, the network attempts to activate into the closest memory (i.e., a training exemplar) of a chemical species encountered in training.

If the element representation consists of core and valence shell electron populations (i.e., inert gas core along with outer s, p, d, and f electrons) and a novel element combination ABC is applied to the input nodes, the network will recurrently converge toward an activation state that minimizes the Euclidean distance between its activation state and that of a training exemplar. If the recursion is initiated with elements A, B, and C fixed, the free stoichiometric nodes vary to match subscripts on the closest isoelectronic species. Therefore, if a net has been exposed to the compound Si_3N_4 in training and the elemental

combination C and N are applied to his recurrent net, activation into the state representing C_3N_4 is highly likely. Hence, such a chemical imagination engine may be used to generate isoelectronic series that may later be evaluated for their thermodynamic plausibility.

1.3.4 Type IV Creativity Machine

The CMs of this class represent various hybridizations of Types I-III. Typically these Creativity Machines consist of many neural network modules. Such cascade architectures may consist of multiple IEs and/ or multiple critic nets. The Creativity Machine chosen for this SBIR is belongs to the Type IV class, and consists of a two preliminary IE nets (i.e., a 'primer' and a Type III recurrent auto-associative net). 14 downstream network modules perform a parallel calculation on the emerging materials concepts, AxByCz, simultaneously calculating accompanying physical and chemical properties.

Using such a Type IV Creativity Machine a broad survey was conducted in a search of new ultrahard binary systems (Thaler, 1998). That work served as a template for the work described below in which a much broader suite of materials properties were projected.

1.4 Materials Properties Sought by the Materials Discovery Creativity Machine

In the current Creativity Machine effort, rather than seek fundamental constitutive parameters, such as dielectric constants and band gaps, for each hypothetical material, emphasis was instead placed upon basic properties known to have more directly measurable and pragmatic thin film engineering value. The properties deemed most useful and known to offer extensive databases for neural network training, were as follows.

1.4.1 Standard Gibbs Free Energy of Formation

As discussed above, the candidate chemical systems generated by the imagination engine are actually isoelectronic variations on the known compounds used as training exemplars. Therefore, additional thermodynamic information must be used to establish the ultimate stability and hence plausibility of any of the machine-proposed materials. For this purpose, the Gibb's standard free energy of formation ΔG_f^o, was calculated for each species recommended by the imagination engine.

1.4.2 Partial Charge

The imagination engine tends to capture the overall patterns of stable chemical species, most notably the tendency of binding atoms to form electronic structures similar to those of inert gases. Therefore, the imagination engine tends to occasionally generate the formulas of ions (i..e., OH^-, $H3O^+$). To ultimately distinguish neutrals within the

resulting database, a separate calculation is needed to provide the most likely charge on the resulting material.

1.4.3 Lattice Constants

An often encountered problem in epitaxial depositions, as well as in the design of multilayer thin film structures, is the inherent mismatch between layers and the underlying substrate materials. The ability to adjust the atomic registry between such layers, using the resulting design database, should prove beneficial. Furthermore, an approximate idea of unit cell dimensions and angles yields some notion of the phases available to each composition.

1.4.4 Melting Point

Often, the intended coating or thin film device must retain its stability and functionality within harsh thermal environments. Having the freedom to choose among myriad materials having the same desired properties, yet a wide spread in melting point, should likewise yield optimized thin film designs.

1.4.5 Density

Since large variations in constituent film densities, may result in large acoustic impedance mismatches between layers, and hence heightened susceptibilty to fracture damage, it is likewise beneficial to have the design latitude to pick matching mass densities. Furthermore, having estimates of bulk material densities and x-ray densities, it is possible to calculate approximate porosities for the resulting films.

1.4.6 Mohs Scale Hardness

Although hardness and bulk modulus do not completely determine wear resistance, they are generally rough guides to thin film durability. Therefore, it is likewise beneficial to have access to these numerical estimates, especially in the context of outer, environmental layers within multilayer structures.

1.4.7 Index of Refraction

Most thin film device codes require specifications for both refractive index and physical thickness of the prescribed film cycles. With the resulting database, a continuum of selections exist for each layer material. Further, the thin film designer has at their disposal numerous choices for the outer impedance matching layer to the external environment (i.e., air, vacuum, or water).

1.4.8 Additional Thermodynamics

Complementary thermodynamic attributes to Gibbs standard free energy of formation, provide consistency checks on their respective values. Heat capacity values may likewise

prove valuable in minimizing thermal shock to delicate multilayer structures or in heat dissipation applications. Furthermore, availability of both ΔH AND ΔS data, allows the estimation of the temperature dependence of Standard Free Energy of Formation through the well known thermodynamic relationship $\Delta G = \Delta H - T\Delta S$. Hence we may assess the stability of any potential chemical system over a wide range of temperatures.

1.4.9 Superconducting Critical Temperature

With the widespread interest in thin superconductivity, it may be quite advantageous to anticipate critical temperatures of single film layers and bulk materials. If models exist, relating the performance of constituent materials to that of superconducting hetero-structures, significant benefit will be seen. Further in satellite and spacecraft applications, it will be useful to have a large repertoire of electronic materials known to be superconductive within the low background temperatures of prescribed mission environments.

1.4.10 Magnetic Susceptibility

Thin film magnets are finding diverse applications in areas ranging from highly efficient supermotors and actuators to the preparation of artificial dielectrics and magnetics for rf, microwave, and low observables applications.

1.4.11 Electrical Conductivity

Foreknowledge of electrical resistance may prove useful for evaporated traces in VLSI design. If bulk chemistry is achievable for the proposed chemical species, lower resistivities may be a major boon to power distribution efforts.

1.4.12 Thermal Conductivity

With the race on in the semiconductor industry for highly thermally conductive substrates and heat sinks, it will be invaluable to have a compilation of many candidate materials rivaling diamond in their heat dissipation characteristics.

1.4.13 Coefficient of Thermal Expansion

For alternating thin film structures exposed to significant excursions in thermal environment, it is highly desirable to minimize strain through the prudent matching of coefficients of thermal expansion. Furthermore, to design optical components requiring a high degree of stability, such as the HR stacks forming laser cavity end mirrors, minimal CTE is desired.

2. Database Generation

The raw data used for the training of the Creativity Machine's imagination engine drew primarily from (1) The TAPP database, (2) Pearson's Tables, (3) The CRC Handbook,

and (4) the text <u>Inorganic Chemistry</u> by Cotton and Wilkinson. Training data for the CM's hetero-associative modules drew heavily upon CRC Handbooks. Many scattered literature sources, including those listed in section 7, were also enlisted for this effort.

2.1 Data Processing and Preliminary Procedures in the Construction of the Creativity Machine

2.1.1 Data Acquisition

Over a period of two years, data has been converted from existing magnetic records and printed texts. In both cases databases were transferred to MS Excel spreadsheets and combined to serve as network training exemplars. Within the course of this Phase I effort a significant fraction of this logging activity consisted of deducing isoelectronic series of the form AxByCz from the classic texts on inorganic chemistry (i.e., Cotton and Wilkinson, 1972).

2.1.2 Conversion to Proper Representation

Once the required chemical databases had been converted to the Excel format, VBA macros were written to convert chemical formulas and data into representations conducive to neural network training. Many experiments were conducted in an attempt to optimize network performance via formula representation. In the context of the imagination engine's large auto-associative map, best results were obtained using an all binary coded representation wherein individual elements were depicted as a vector containing binary coded row number (from the periodic table) as well as binary coded electron populations for ground state valence s, p, d, and f shells (e.g., hydrogen was encoded as the super-vector, $\{(0,0,1), (0,1), (0,0,0), (0,0,0,0), (0,0,0,0)\}$, where the vector $(0,0,1)$ represents the first period of the periodic table, $(0,1)$ represents a single s-electron, and subsequent vectors $(0,0,0,0)$ represent empty d and f electron shells.). Stoichiometric subscripts were likewise binary coded. Similarly, it was found that the most efficient and economical formula representation for the hetero-associative modules consisted of analog numbers corresponding to the inert gas cores of each constituent element, along with decimal representations of valence s, p, d, and f electron populations. Stoichiometries were encoded as decimal, analog values for these hetero-associative modules (e.g., neon would be encoded as the vector $\{2, 2, 6, 0, 0\}$, indicating a helium core of $Z=2$, followed by 2 s, 6 p, 0 d, and 0 f electrons).

2.1.3 Auto-associative Filtering of Training Data

The lumped materials database of over 40,000 chemical exemplars, was used to train a filtering auto-associative network. The utility of such a network is the ability to detect outliers and anomalies within a database (Thaler & Conrad, 1998). In retrospect, many of the data anomalies detected via auto-associative filtering turned out to be nothing more than typographical errors. Other anomalies that could not be attributed to transcription error, were retained and were assumed to represent real deviations from overall materials trends.

2.1.4 Symmetrization of Training Data

Cumulative experience in training both auto- and hetero-associative networks in the context of materials properties prediction taught that the ordering of elements applied to the inputs of the net had a significant impact upon network prediction error. Therefore, if appropriate precautions were not taken in the preprocessing of the training exemplars, significantly different calculated properties would result for the formulas AxBy and ByAx applied to the same trained feedforward network. Therefore, in the case of binary compounds, the training exemplars were doubled by reversing the elements and stoichiometries. For ternary systems, the training exemplars were expanded six-fold, by the inclusion of orders ABC, BCA, CAB, ACB, BAC, and CBA.

2.1.5 Network Optimizations

Hidden layer architectures were optimized in terms of RMS test prediction error. Genetic algorithms were used to vary layers, hidden units, learning rate, momentum, and annealing noise toward optimal values that minimized network generalization error.

2.1.6 Conversion to Spreadsheet-Implemented Networks

Rather than rely upon linear programming languages such as C, the neural networks employed in this study were implemented by 'weaving' together cells within a spreadsheet application via relative referencing (Thaler, 1996e, 1998a,b,c). Such spreadsheet-implemented networks are transportable and object-oriented so that they may be rapidly positioned and mated with other neural networks within complex cascade structures. This capacity to rapidly implement and test such parallel architectures greatly accelerated the prototyping, testing, and final runs of the materials discovery Creativity Machine produced in this effort.

2.1.7 Conversion to Active Server Pages

For each neural network implemented as a woven spreadsheet net, a corresponding VisualBasic Script version was created for integration within an Active Server page (ASP). These were and are currently being used in a complementary role to the completed Creativity Machine, in providing numerical estimates of both chemical and physical properties of elements, and both binary and ternary compounds within and beyond the bounds of the dynamic database generated by the materials discovery CM. These networks are available at the joint IEI-AFRL CRDA site at http://www.imagination-engines.com/storefront/crda.htm.

2.1.8 Connection of Network Modules

Following the training and refinement of each ANN, the corresponding spreadsheet module was generated and then pasted within an individual spreadsheet of an Excel

workbook. The network cascade was linked together by threading network inputs and outputs between the individual worksheets of the Excel workbook (Fig. 6).

2.1.9 Governing Macros

A controlling algorithm was created to drive the CM cascade as well as to govern the archiving of discovery data. Summarily the algorithm could be placed in either a binary or ternary system search mode wherein elements A, B, and C (within the ternary search) were cycled through and then used to pin the element inputs of the IE net as stoichiometric nodes were stochastically activated. Each new potential chemical formula was then distributed to all of the hetero-associative properties nets to calculated anticipated property values.

2.1.10 Conversion to MS Access Database

Both cumulative binary and ternary systems databases were imported from MS Excel to MS Access to implement SQL queries. Further, this database was connected via an ODBC model, to the World Wide Web for public access. The HTML interface is located at http://www.imagination-engines.com/af98-190/password.asp (password entry only).

2.2 Materials Discovery Creativity Machine (MDCM) Architecture
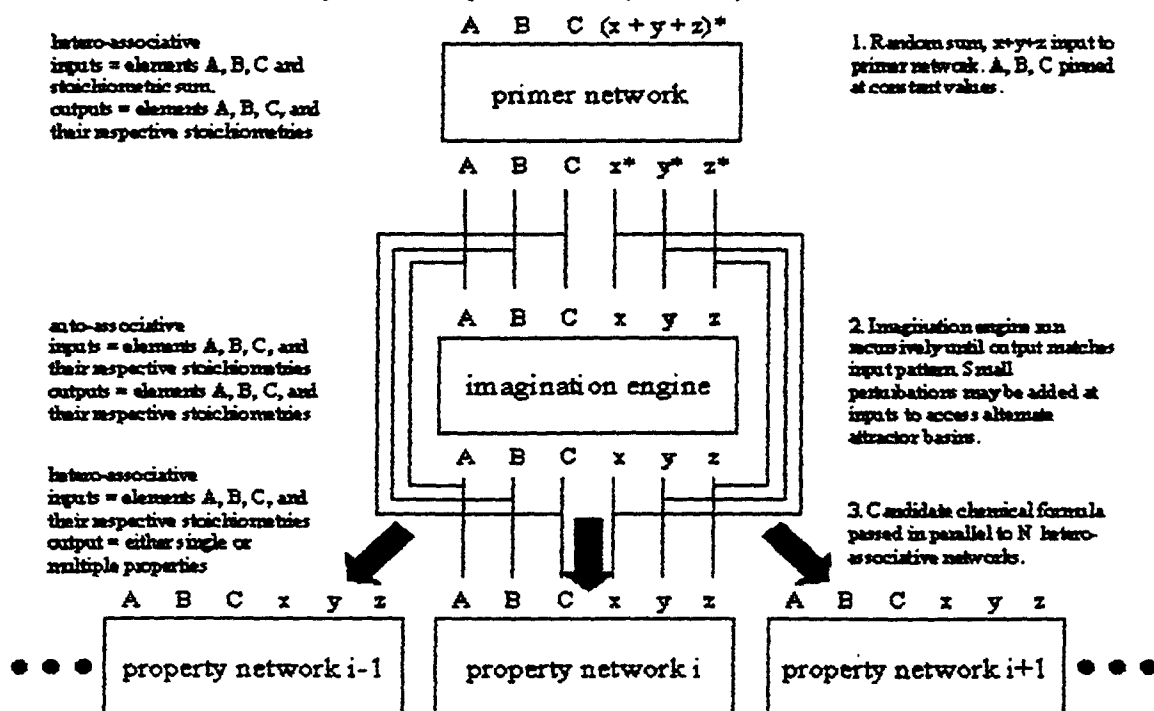

Figure 5A. Highly schematic diagram of the Materials Discovery Creativity Machine. A primer net first recommends an approximate stoichiometry. A type III imagination engine then gravitates toward some exact stoichiometry. A system of parallel hetero-associative modules then estimates accompanying physical and chemical properties.

The overall architecture of the Materials Discovery Creativity Machine is shown in Figure 5A. In the first stage of the calculation, the governing algorithm generates the electron representations of three separate elements, A, B, and C, using the analog representation. Furthermore, a random number is generated, representing the sum of stoichiometric indices, x+y+z. the primer net then calculates the most likely distribution of this sum into stoichiometric indices x, y, and z. If the sum x+y+z does not involve the sum of plausible stoichiometries, the network outputs settle into the closest achievable stoichiometry (i.e., for input of A="H" and B="O" and stoichiometric sum=2, this net would yield stoichiometric outputs of x=2 and y=1, representing H_2O. Likewise, a stoichiometric sum of 3 would yield H_2O, but a sum, x+y+z, of 4 would yield H_2O_2.)

The formula, represented at the primer net's outputs, is then converted to the full binary coded representation. These BCD elements and stoichiometries now serve as inputs to the imagination engine. Within the first recurrent pass through this network, internal synaptic noise is set at a mean value of 0.06, consistent with the prescription of Equation 1. The auto-associative IE is then run recurrently, repeatedly circulating outputs back to inputs (with A, B, and C fixed). With each successive pass through this network, the mean synaptic noise is modulated by the current delta error between the output and input vectors. Should this recurrent net stochastically locate a sufficiently deep attractor basin (representing either a stoichiometric memory, or a generalization of its memories), the delta error between output and input patterns will converge toward a small number. At this point, the activation state of the IE stabilizes and there is no further significant network evolution. Once this network convergence is achieved, the complete formula, AxByCz, represented at the network's outputs, is now relayed in parallel to the individual property networks, following conversion back to the analog representations for both elements and stoichiometries. The governing algorithm then archives the potential chemical formula, along with accompanying calculated property values.

In this initial effort, the governing algorithm generated compositions consisting of elements A, B, and C, in alphabetical order with A≠B≠C. To achieve this end, a series of three nested FOR loops were constructed within the VBA macro, corresponding to each of the elements A, B, and C, respectively. The B loop began with the next alphabetical choice of element symbols following A. Likewise, the C loop began with the next alphabetical element symbol entry following that of B.

For each combination of elements ABC generated by these nested loops, ten trials were conducted, experimenting with the stoichiometries x, y, and z. In the early stage s of this Creativity Machine, stoichiometries were restricted to values less than or equal to 7. Recursion through the IE net continued until the RMS delta error between inputs and outputs fell below the predetermined threshold of 0.05

Weaving of the network cascade was achieved by embedding the individual ANN modules depicted above in Figure 6, into successive worksheets within an Excel workbook. Relative referencing between individual sheets served as a data bus,

conveying the calculated formula emerging from the primer and IE nets to the downstream property nets.

Primer

$$\blacktriangledown$$

Imagination Engine

$$\blacktriangledown$$

$$A_xB_yC_z$$

ΔG_f^o	q	a,b,c	α, β, γ

T_{melt}	ρ_m	Mohs	n_D

ΔH_f^o	ΔS_f^o	T_c	χ_m

ρ_e	κ_T	α_T	MW
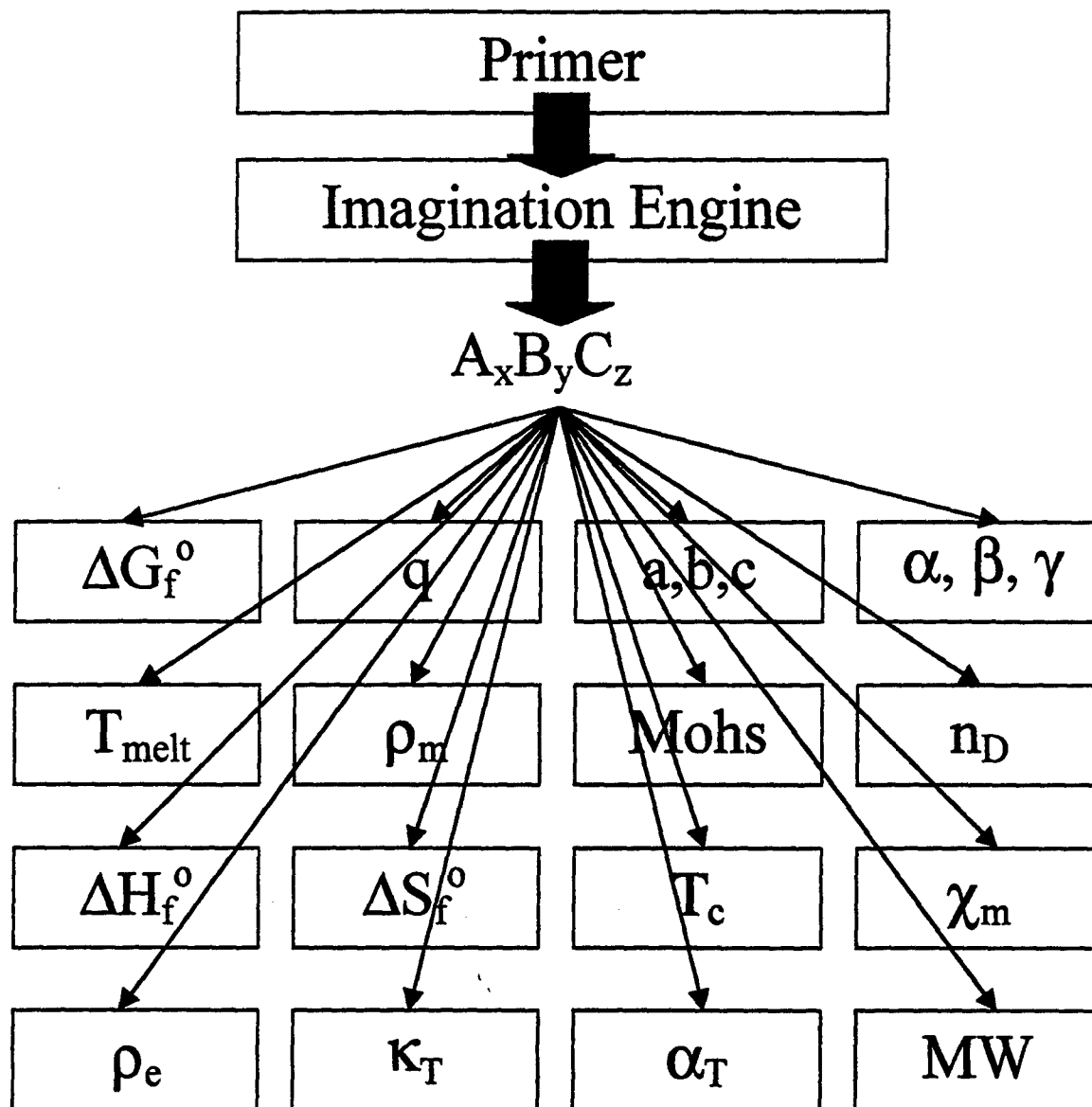
Figure 5B. Highly schematic diagram of the Materials Discovery Creativity Machine. A hypothetical formula of the form AxByCz originates within the primer-imagination engine stage of the Creativity Machine cascade. Physical and chemical properties are then estimated in parallel calculations within the fifteen hetero-associative modules and a single algorithmic macro that calculates exact molecular weight, MW. ΔG_f^o is Gibb's Free Energy of Formation, q is the charge on hypothetical species, a, b, c, α, β, γ are lattice parameters, T_{melt} is melting point, ρ_m is the mass density, Mohs is Mohs Scale hardness, n_D is the sodium D-line refractive index, ΔH_f^o is the standard enthalpy of formation, ΔS_f^o is the standard enthalpy of formation, Tc is the superconducting critical temperature, χ_m is the magnetic susceptibiliy, ρ_e is electrical resistivity, κ_T is thermal conductivity, α_T is coefficient of thermal expansion, and MW is the molecular or formula weight.
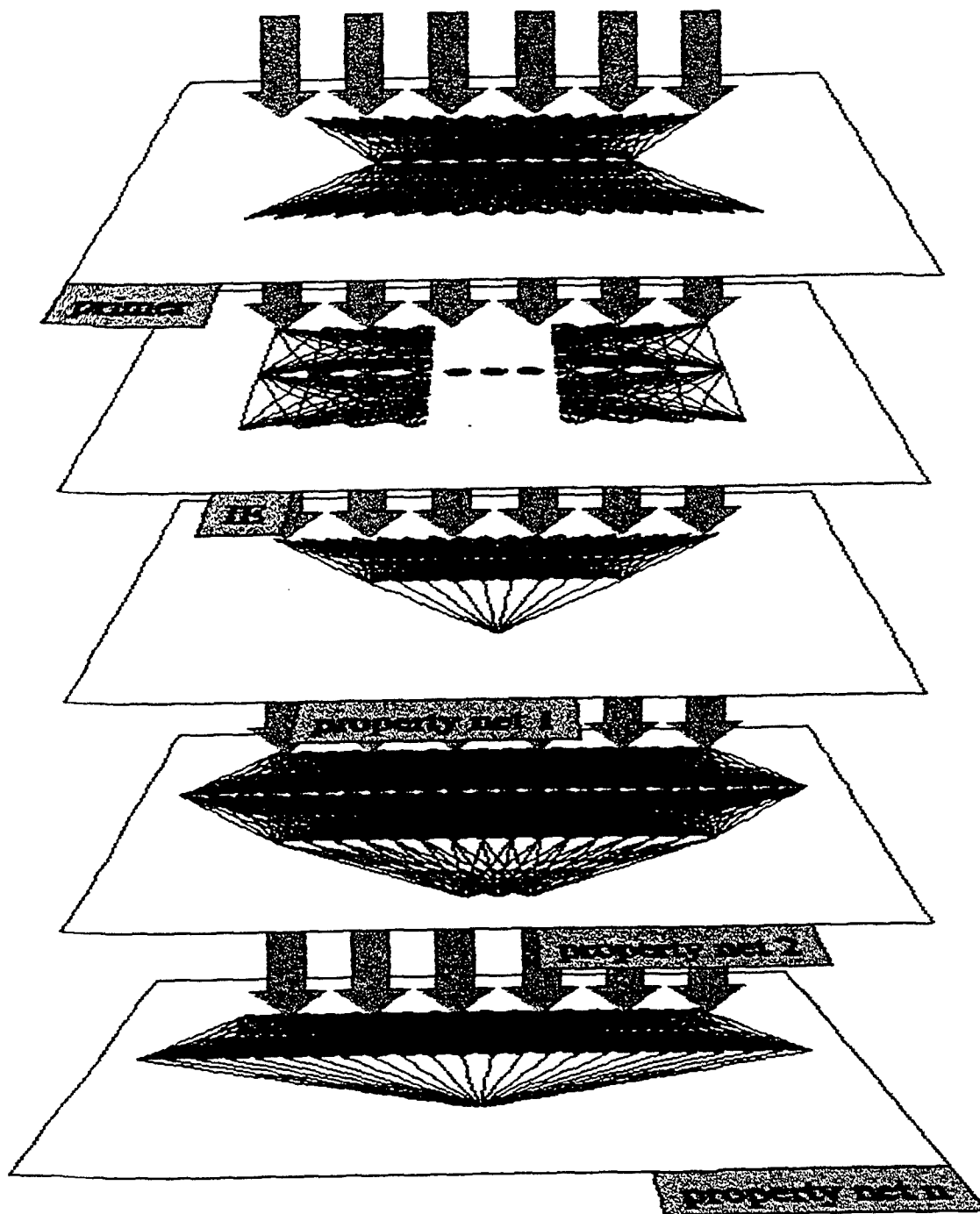
Figure 6. Implementation of Materials Discovery Creativity Machine within Excel Workbook. Each worksheet, corresponding to a tabbed layer within the diagram, contains a single neural network module, knitted by relative reference between individual spreadsheet cells. Likewise, the individual modules are linked between worksheets via relative reference. Between the auto-associative IE and lower hetero-associative nets, the representation is converted from full BCD to base 10 analog format.

2.2.1 Primer Network

A three layer net having 16 input, 10 hidden layer, and 18 output nodes was trained upon roughly 15,000 binary and ternary systems to map inputs consisting of elements A, B, C, and a total stoichiometry (x+y+z) to identical outputs A, B. C along with the most likely analog stoichiometric indices, x, y, and z. Essentially, this network was designed to accelerate the performance of the IE net in seeking its attractor basins by supplying initial analog estimates of stoichiometry. Because the stoichiometric sum could be parametrically increased, multiple stoichiometries could be established for any elemental combination A, B, and C. Thus for A=H, B=O, and C void, an input of x+y+z = 3 to this net would output x=2, y=1, z=0. Similarly, increasing the input quantity x+y+z to 4 would output the values x=2, y=2, z=0. In other words, the primer net may be thought of as distributing stoichiometry among the constituent elments of a system A, B, and C. The additional outputs of elements A, B, and C is a crosscheck as to whether the primer perceptron is in fact sensing the correct elements at its inputs. Therefore the appearance of any elements at the outputs, other than the input triplet, would invalidate the recommended stoichiometric distribution x, y, and z.

The 15,000 chemical systems utilized in training this net were drawn at random from the combined TAPP database and Pearson's Tables. 20% of this data was reserved for network testing, wherein RMS prediction error was minimized to approximately 0.1 stoichiometric unit. Following testing, the reserved data was recombined with the training data and network training resumed until training error likewise reached 0.1 stoichiometric unit

2.2.2 Imagination Engine

The largest network within the cascade was the auto-associative memory constituting the imagination engine. Training data for this network consisted of roughly 15,000 entries from the TAPP database, 16,000 from the Pearson's Tables, 6,000 retrieved from scattered literature, and approximately 3,000 exemplars 'manually' generated from Cotton and Wilkinson's (1972) classic text <u>Advanced Inorganic Chemistry</u>. In the case of the latter source, whole families of chemical systems could be either recovered or deduced from chapters and sections devoted to related compounds and systems. Therefore, the spans of isoelectronic series could be established (i.e., Ax By, A'x B'y, are known to form, but Ax''B''y does not, where A, A', A'' possess identical valence shells, but differing atomic cores).

The representation used for training this network consisted of chemical formulas cast into entirely binary coded form. Each of the elements was represented by 16 bits, while each stoichiometric index was indicated through 8 bits. Full compound or system representation consisted of the elements A, B, and C, followed by the stoichiometries, x, y, and z. Therefore, the total number of inputs (or outputs) to this auto-associative network totaled 72. Training was continued until RMS training error fell below 0.1% for each of the Boolean output nodes.

2.2.3 Standard Free Energy of Formation Network

A three layer network, consisting of 18 inputs, 11 hidden layer units, and a single output node, was trained to within 6% rms (~ 30 kcal/mole) error for the total free energy range of the training exemplars, -450 to +50 kcal/mole. Inputs consisted of the electronic representation of the 3 element compound and accompanying stoichiometries x, y and z.

2.2.4 Partial Charge Network

A three layer feedforward network, having 18 inputs, 15 hidden layer units, and a single output node, was trained to 3% rms of its total output range of 9 charge units. The training set consisted of approximately 3,000 exemplars, approximately 800 of which were known ionic species. The remainder of the training set was made up of known neutrals borrowed from the 10,000 compounds used for stoichiometry training. Inputs to the net consisted of the valence shell configurations of each constituent element, A, B, and C, along with the accompanying x, y, and z stoichiometries.

2.2.5 Lattice Constants Network

Unit cell lengths and bond angles were calculated via a three layer network having 21 input, 60 hidden layer, and 6 output units. Both elemental and stoichiometric inputs to this network were binary coded. The training set for this net consisted of 5,300 exemplars randomly selected from TAPP and Pearson databases. RMS testing error for this network was 10% of the range for each of the lattice constants (approximately 1 Angstrom for lattice constants and 2 degrees for unit cell angles).

2.2.6 Melting Point Network

This four layer net consisted of 18 input, 29 first hidden layer, 18 second hidden layer, and 1 output node. 9,000 training exemplars were randomly drawn from the combined TAPP, CRC, and Pearson databases. 20% of these were randomly selected as a reserved test set. Test accuracy of this network was reduced to 5% of the range of melting temperatures of 4000 degrees C, or approximately 200 degrees C. Note that melting point estimates from this net only become useful at the higher melting points, where an error of 200 degrees C is only a small fraction of the calculated temperature.

2.2.7 Density Network

This three layer network, capable of handling quarternary systems, consisted of 24 inputs, 25 hidden layer units, and a single output representing estimated mass density. Approximately, 15,000 training exemplars were used, drawn randomly from the combined TAPP, CRC. and Pearson databases. 20 % of these exemplars were reserved for test purposes. The overall RMS prediction error for this net was reduced to 5% of the overall range of density values, 36, for an average error of 1.8 grams per cubic centimeter.

2.2.8 Mohs Scale Hardness Network

A three layer network) was trained to within 10% rms error (1 Mohs Scale unit) in mapping binary compounds to their anticipated Mohs Scale hardness. In all, 339 training exemplars were used in training this net, chosen largely from minerological references. If for any given compound within the net's training set, there arose a choice of phase, the hardest of the room temperature and atmospheric pressure phases was chosen. Therefore, within this training set, carbon is represented as diamond, and not graphite.

2.2.9 Index of Refraction Network

A four layer network was used to map composition to sodium D-line refractive index. The net consisted of 18 inputs, 26 first hidden layer, and 19 second hidden layer, and a single output representing refractive index. A total of 2,800 training exemplars were collected largely from the CRC Handbook and scattered literature. 20% of these exemplars were reserved as test data. RMS Test prediction error was minimized to 3% of the range of available index values, 3.2, for an average error of approximately 0.1.

2.2.10 Thermodynamics Network

Composition was mapped to Gibb's free energy, enthalpy and entropy of formation, as well as heat capacity at constant pressure via a three layer net having 18 input, 38 hidden layer, and 4 output notes corresponding to the aforementioned thermodynamic quantities. A total of 565 training exemplars were collected from the literature (primarily CRC). 20% of this data was reserved for test purposes. Total RMS prediction error was minimized to 4% of the range of each of these thermodynamic functions, amounting to approximately 32 kcal/mole for free energy and enthalpy, and 2.8 kcal/mole-degK for entropy and heat capacity values.

2.2.11 Superconducting Critical Temperature Network

A three layer net consisting of 18 inputs, 32 hidden layer, and a single output, was used to map composition to approximate superconducting critical temperature. Approximately 4,100 training exemplars were collected from both the CRC Handbook and the NIST Superconductor Database, 20 % of which were reserved for network testing. Overall RMS prediction error was minimized to 10% of the overall range of critical temperatures, 21 deg K, for an average prediction error of 2 deg K.

2.2.12 Magnetic Susceptibility Network

A four layer net was used to map composition to magnetic susceptibility in units of 1E-6 CGS. The net consisted of 18 input, 30 first hidden layer, 16 second hidden layer, and two ouput nodes representing first algebraic sign and the base 10 logarithm of the absolute value of magnetic susceptibility, χ_m. In total, 3,700 compounds and corresponding magnetic susceptibilities were gathered from the literature, 20% of which

were reserved for network testing. Overall RMS prediction error was reduced to 5% of the overall range of $\log(\chi_m)$, 5.6, for an average error of .28 in $\log(\chi_m)$.

2.2.13 Electrical Resistivity

A three layer net was used to map composition to the base 10 logarithm of resistivity, ρ, measured in units of Ω-m. Approximately 1,000 training exemplars were collected from diverse literature that included the CRC Handbook. 20% of this data was reserved for testing. Overall RMS prediction error was minimized to 5% of the range of $\log(\rho)$, 21, for an average error of approximately 1.05. Therefore, resistivity values were accurate to within approximately a factor of 10.

2.2.14 Thermal Conductivity Network

A three layer network was used to map composition to thermal conductivity, κ_T, measured in units of mW/cm-K. This net consisted of 18 input, 8 hidden layer, and a single output node representing thermal conductivity. 1,045 training exemplars were collected from scattered literature and the CRC Handbook. 20% of these exemplars were reserved for testing purposes. The final RMS prediction error attained was 6.5% of the range of thermal conductivity values trained upon, 4.6, yielding an average prediction error of 0.3 mW/cm-K.

2.2.15 Coefficient of Thermal Expansion Network

Coefficient of thermal expansion, α, in units of 1E-6/degK, was calculated using a three layer net having 18 inputs, 10 hidden layer, and a single output node representing coefficient of thermal expansion. A total of 600 training exemplars were collected from diverse literature and the CRC Handbook. 20% of these exemplars were reserved for network testing purposes. Ultimately, RMS test error was minimized to 5% of the total range of CTE values within the training database, 76, yielding an average prediction error of 3.8 E-6 / degK.

2.3 Materials Discovery Creativity Machine (MDCM) Runs

Preliminary to the final run of the completed Materials Discovery Creativity Machine (MDCM), multiple investigations were carried out using only the imagination engine to examine the plausibility of the resulting chemical compositions. The conclusion of this study was that whereas each compound generated by the IE had precedence in terms of its valence shell configuration, not all materials proposed were likely. That is, each hypothetical compound produced by the imagination engine was isoelectronic with some material already encountered among the training exemplars. Therefore, the realizability of each of these chemical systems was plausible only within the context of the concurrently calculated standard Gibb's free energy of formation. Furthermore, any stoichiometric series AxByCz (where A, B, and C are fixed and x, y, and z vary) ranked in ascending order of free energy of formation, showed the most commonly encountered stoichiometries invariably near the top of the list. However, some not so well known

stoichiometries also appeared among the thermodynamically stable candidates, indicating either (1) flaws in the overall system, (2) potentially stable, yet heretofore kinetically inaccessible species that could be achievable via advanced thin film techniques, or (3) mixed stoichiometries, solid solutions, or interstitial substitutions.

Example stoichiometries recovered from an early test run of the imagination engine are shown for the specific cases of both vanadium and iron oxides, as well as for sodium chloride in Table 1. In each of these cases, the stoichiometric combinations discovered by the IE are ranked by calculated Gibb's free energy of formation, from most to least stable. In these cases, and many more studied, the known stoichiometries tend to populate the upper portions of these ordered tabulations. The other proposed materials may require explanation, either as artifacts of the calculation, or kinetically inaccessible. In the case of the oxides, for instance, thin film growth may allow the formation of sub- and super-oxides, depending upon the oxygen background pressures used. Therefore, we can expect a wide variation in thin film oxide stoichiometries manifesting stability.

The variety in the sodium chlorides may require chemical explanation in terms of other known quasi-stable Na_xCl_y stoichiometries that include charge transfer complexes as well as heavy chloride diffusion into NaCl at elevated temperatures that results in F-center defect production. The multiple entries Na_xCl_x are interpreted as variations on the formula NaCl induced by variations in formula and unit cell conventions within the databases mined for neural network training exemplars.

One feature of the imagination engine studied at great lengths, was the frequency with which this network generated various stoichiometries x, y, and z, for fixed elemental composition A, B, and C. These frequency were in turn indicative of the depths of the attractor basins representing these stoichiometries. One case in point is the plot of Figure 7, where first the raw frequency of the Fe_xO_y stoichiometries is shown, alongside a similar plot wherein each frequency is modulated by the simultaneously calculated standard free energy of formation. We see that in the latter, the attractor landscape clearly yields the known iron oxide stoichiometies of FeO, Fe_2O_3, and Fe_3O_4. The iron-enriched species Fe_5O_4 is puzzling and may require further explanation.

Certainly, these frequencies would be necessary to the most thorough tabulation of stoichiometries. However, since a single pass through all elemental combinations requires days of CPU time, such a study would require months to complete. In the meantime, at least a portion of the novel stoichiometries could be no more than spurious, shallow attractor basins of the network.

Based upon theses observations, we may relaibly state that the stoichiometries predicted by the imagination may be viewed as probabilistically correct. Those chemical formulas generated by the imagination engine having large negative free energies of formation are very likely to represent plausible chemical systems. To thoroughly assess plausibility, many thousand iterations would be needed for a given composition A, B, and C, varying x, y, and z to establish the relative depths of the attractor basins involved.

Table 1. Example Stoichiometries from the Imagination Engine

| IE	Known	ΔGf	IE	Known	ΔGf	IE	Known	ΔGf
V2O3	yes	-547.1	Fe4O5		-544.3	NaCl2		-540.5
V5O7		-546.9	Fe5O6		-544.3	Na3Cl4		-525.3
V4O7		-546.8	FeO	yes	-543.8	Na5Cl5	yes	-503.1
VO2	yes	-545.8	Fe6O7		-543.3	Na6Cl6	yes	-446.7
V6O7		-542.2	Fe3O5		-542.6	Na7Cl2		-339.9
VO3	yes	-535.4	Fe2O3	yes	-542.2	NaCl4		-263.7
V2O7		-533.9	Fe7O5		-536.6	Na2Cl2	yes	-141.1
VO	yes	-533.5	Fe3O4	yes	-535.4	NaCl	yes	-83.1
V2O5	yes	-520.6	FeO2		-531.1	.		
			.			.		
V7O		-344.7	FeO4		-363.2	.		
.			.					
V5O		-135.0	Fe6O		-42.1	.		
V3O2		-79.7	Fe2O		-37.7	.		
V2O		-19.0	Fe3O		-21.2	.		
V4O		+4.7	Fe5O		-18.8	.		
V3O		+20.5	Fe4O		-15.3	Na4Cl		-12.5
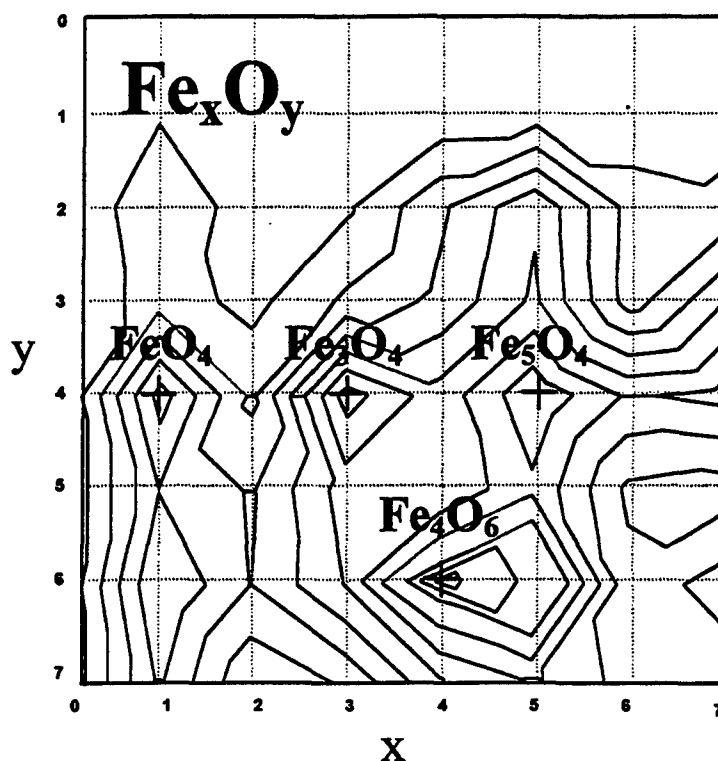
*Figure 7. Attractor Basin Lanscape of the Imagination Engine for the Iron Oxides.
Closed contours represent minima.*

2.3.1 Binary Systems

In the first run of the complete materials discovery Creativity Machine, the binary systems AxBy were surveyed. To contact with existing literature, the stoichiometries x and y were restricted to values less than or equal to 7. Approximately 70,000 chemical systems and accompanying properties were amassed within 24 hours using a Pentium II processor running at 200 MHz. This Excel database was then converted into an Access database in preparation for making it accessible to queries via the internet.

2.3.2 Ternary Systems

In the second run of the materials discovery Creativity Machine, a survey was made of ternary systems AxByCz, again pinning the stoichiometries x, y, and z to values less than or equal to 7. Following 72 hours of sustained parallel runs on two Pentium II processors, a total of approximately 160,000 ternary systems and their accompanying properties were archived within an Excel database. This data was then converted into an Access database in preparation for internet queries.

3. The Resulting Database

Using the Database – The materials database generated by the Creativity Machine are accessible through an html interface that supports standard query language (SQL). The password gateway to this interface is at http://www.imagination-engines.com/af98-190/password.asp (i.e, passwords are obtained on a membership basis, to obtain a password contact sthaler@ix.netcom.com). Because an operator may intersect regimes of multiple materials properties, this database, coined the "MatterHorn" forms an effective materials design tool that may be used to fulfill broad materials specifications. Therefore, it is possible, for example, to interrogate either binary or ternary databases for materials having a superconducting critical tempertaure above 20 deg K, while also manifesting a Knoop Hardness of in excess of 9.0, ordering by Tc, etc.

As elaborated upon in section 2.3, this database is to be used in a probabilistic sense, evaluating the plausibility of any proposed stoichiometry in terms of the projected free energy of formation. Further, materials properties estimates are to be regarded as only approximations and not precise values gleaned through meticulous laboratory procedures. Nevertheless, this database does represent a broad brushstroke survey of all binary and ternary chemistry. Furthermore, it is a chemical compendium that is based largely upon composition, considering in the majority of its records, the physical and chemical properties of an 'average' phase of any given stoichiometry AxByCz.

3.1 Use of the MatterHorn Database

Both of the MatterHorn databases (binary and ternary systems) consist of the same materials properties headers, corresponding to the symbols shown in Table 2. All quantities defined there should be self-evident to the materials scientist, with the

exception of the delta column, corresponding to the convergence error (final difference of the IE's input and output patterns). This quantity is included as a diagnostic check that the imagination engines did indeed relax into a network attractor basin and that the stiochiometry is not the result of some unforeseen computational pathology.

Table 2. Attributes incorporated into the MatterHorn Database

HEADER	MEANING	UNITS
DELTA	RESIDUAL INPUT-OUTPUT ERROR IN IE NETWORK	dimensionless
A, B, C	ELEMENT SYMBOLS	text
x, y, z	RESPECTIVE STOICHIOMETRIC INDICES FOR A, B, & C	dimensionless
MW	MOLECULAR OR FORMULA WEIGHT	grams
$DGF = \Delta G_f^o$	STANDARD GIBBS FREE ENERGY OF FORMATION	kcal/mole at 25 deg C
CHG	PARTIAL CHARGE ON HYPOTHETICAL CPD/SYSTEM	electron charge
CELL A,B,C	LATTICE CONSTANTS	Angstroms
CELL ALPHA, BETA, GAMMA	UNIT CELL ANGLES	degrees
MP	MELTING POINT	deg K
DENS	DENSITY	grams/cm^3
MOHS	MOHS SCALE HARDNESS	dimensionless
INDEX	INDEX OF REFRACTION	dimensionless
$DH = \Delta H$	ENTHALPY OF FORMATION	kcal/mole at 25 deg C
S	ENTROPY OF FORMATION	cal/deg-mole at 25 deg C
$CP = C_p$	HEAT CAPACITY AT CONSTANT PRESSURE	cal/deg-mole at 25 deg C
$TC = T_c$	SUPERCONDUCTING CRITICAL TEMPERATURE	deg K
CHI	MAGNETIC SUSCEPTIBILITY	10E-6 CGS
RHO	ELECTRICAL CONDUCTIVITY	ohm-m
TCON	THERMAL CONDUCTIVITY	mW/cm-K at 300 deg K
CTE	COEFFICIENT OF THERMAL EXPANSION	10E-6 /deg K at 300 deg K

To query this database, use standard SQL commands of the form,

Select * from [database] **WHERE** [logical clause] **ORDER BY** [some criterion].

Once a databse is selected using the binary or ternary buttons, the base command Select * from [database] is automatically generated within the text area field of the html form. Simply add the required clause(s) to this query root.

Below are listed some example queries that may be executed against these respective databases:

```
Select * from BINARIES WHERE A='Al' AND B='O'
Select * from BINARIES WHERE TC>20 AND RHO>1000
Select * from BINARIES WHERE INDEX>3 ORDER BY INDEX
Select * from BINARIES WHERE CHI>50000 AND DENS>15
Select * from BINARIES WHERE TC>19 ORDER BY TC
Select * from BINARIES WHERE MOHS>9.5 AND DENS<5
Select * from TERNARIES WHERE A='Al' AND B='O' AND C='Si' ORDER BY RHO
Select * from TERNARIES WHERE TC>20 AND RHO>1000
Select * from TERNARIES WHERE INDEX>3 ORDER BY INDEX
Select * from TERNARIES WHERE CHI>50000 AND DENS>15
Select * from TERNARIES WHERE TC>19 ORDER BY TC
Select * from TERNARIES WHERE MOHS>9.5 AND DENS<5
```

In performing queries of both the binary and ternary databases, it is important to remember that the current version of the underlying active server script can only deal with compositions ABC, where A, B, and C appear in alphabetical order. Therefore a query of the form, **Select * from TERNARIES WHERE A='V' AND B='Nb' AND C='O'** would return no results, since the constituent elements have not been entered in the correct ascending order. The correct sequence for this query would read **Select * from TERNARIES WHERE A='Nb' AND B='O' AND C='V'.**

Likewise, it is cautioned that bandwidth limitations imposed by the internet, prohibit large samplings of the MatterHorn database. It is therefore advised that enough restrictions be placed on the query so as to limit the potential number of hits. Otherwise, the server script will time out after approximately two minutes, followed by the corresponding error message. Therefore, for intensive queries of the SBIR database, it is recommended that the MS ACCESS database be directly interrogated.

Furthermore, as discussed below, it is likewise recommended that the user stipulate an ordering by the standard Gibb's free energy of formation (DGF) to rank the resulting query results in terms of thermodynamic plausibility. In this case, the query would be appended with a clause of the form ...**ORDER BY DGF.**

Figures 8 and 9 show successive stages of the on-line query. Presuming that a database member has successfully logged-in, they are presented with a text area form for the entry of their query. The html interface is initialized in the binary system search mode, but may be placed in the ternary search mode (Figure 8) through the button marked **new ternary query**. The user may return to the binary search mode by pushing the **new binary search** button, at any time.

The query stem is automatically generated within the text area form element and will read either (1) **Select * from BINARIES** or (2) **Select * from TERNARIES**, depending upon the search mode. The user must add the correct clause to this stem, typically beginning with the word **WHERE**, followed by the search constraints. Note that the query has been preserved at the SQL level to afford the user the greatest flexibility in searching these discovery databases. Once the query line is completed, the user pushes the **Submit Query** button. It is highly recommended that the display window be enlarged beyond 640 pixels to accommodate the wide tables used to display query results (Figure 9).

Address: http://www.imagination-engines.com/af98-190/terrquery.asp

MATTERHORN

a theoretical database of potential binary and ternary chemical systems.

The MatterHorn is a collaborative effort between Imagination Engines, Inc. and the Air Force Research Laboratory Materials and Manufacturing Directorate (AFRL/ML) Manufacturing Technology Division, Materials Process Design Branch (AFRL/MLMR), Wright-Patterson AFB, OH

new binary query

new ternary query

HEADER	MEANING	UNITS
DELTA	RESIDUAL INPUT-OUTPUT ERROR IN THE NETWORK	dimensionless
A, B, C	ELEMENT SYMBOLS	text
x, y, z	RESPECTIVE STOICHIOMETRIC INDICES FOR A, B, A C	dimensionless
MW	MOLECULAR OR FORMULA WEIGHT	grams
$DGF = D_3^o$	STANDARD GIBBS FREE ENERGY OF FORMATION	kcal/mole at 25 deg C
CHG	PARTIAL CHARGE ON HYPOTHETICAL CPD/SYSTEM	electron charge
CELL A,B,C	LATTICE CONSTANTS	angstroms
CELL ALPHA, BETA, GAMMA	UNIT CELL ANGLES	degrees
MP	MELTING POINT	deg K
DENS	DENSITY	grams/cm^3
MOHS	MOHS SCALE HARDNESS	dimensionless
INDEX	INDEX OF REFRACTION	dimensionless
$DH = \Delta H$	ENTHALPY OF FORMATION	kcal/mole at 25 deg C
S	ENTROPY OF FORMATION	cal/deg-mole at 25 deg C
$CP = C_p$	HEAT CAPACITY AT CONSTANT PRESSURE	cal/deg-mole at 25 deg C
$TC = T_c$	SUPERCONDUCTING CRITICAL TEMPERATURE	deg K
CHI	MAGNETIC SUSCEPTIBILITY	10E-6 CGS
RHO	ELECTRICAL CONDUCTIVITY	ohm-m
TCON	THERMAL CONDUCTIVITY	mW/cm-K at 300 deg K
CTE	COEFFICIENT OF THERMAL EXPANSION	10E-6 /deg K at 300 deg K

Query string to be executed against TERNARIES

Select * from TERNARIES WHERE TC >20 ORDER BY TC

Submit Query

Figure 8. HTML Interface to the MatterHorn Database. Queries are entered into the text area form in standard query language. The submit button launches the database query.

Your query: Select * from TERNARIES WHERE TC >20 ORDER BY TC

Here are your query results:

Figure 9. Dynamically Generated HTML in Response to Query Shown in Figure 8.

4. Conclusions

4.1 Neural Network Methodology

The guiding philosophy behind the materials discovery Creativity Machine constructed as part of this Phase I SBIR effort mirrors nature, in that once a combination of elements A, B, and C are specified, the possible ranges of allowable stoichiometries fall within a discrete manifold of possibilities. In turn, these resulting compositions AxByCz then restrict the kinds of potential phases accessible. Considering phase to be a function of stoichiometry, approximate physical and chemical properties may be estimated and associated with any given formula AxByCz. Final analysis of the prediction accuracy offered by the downstream hetero-associative property modules reveals that the calculation error of these nets typically is not sufficiently small enough to resolve the spread in property values between separate phases of any given composition. Nevertheless, the coarse survey of all binary and ternary chemistry offered by this materials discovery Creativity Machine is unprecedented in its scope, potentially offering a boon to both thin-film and bulk materials discovery efforts.

Further, in overview, the computational materials search accomplished during this study minimally has value as a broad survey of how materials properties vary as a function of elemental substitution, A, B, and C. Optimally, however, the machine-manufactured databases produced offer an accurate prediction of potential stoichiometries, AxByCz, and phases delivering a wide range of technologically useful characteristics. Perhaps the reality lies somewhere between these two extremes.

Weaknesses in the methodology that have detracted from the overall accuracy of this discovery system, include:

(1) The wide variations an inherent errors that exist within the historical materials databases – Oftentimes different sources quote widely varying property values for a given composition and phase. These discrepancies certainly point to some degree of error within the empirical materials property measurements and most assuredly impact the accuracy of the neural network fit to that data.

(2) Sparseness of data – Needless to say, the literature contains vast gaps in materials properties data, for a variety of reasons. In many cases materials have been difficult or dangerous to synthesize (i.e., those containing radioactive isotopes). In other cases, either the lack of commercial utility or natural abundance has dissuaded materials characterization. However, in many cases, the shear number of combinatorial possibilities (i.e., the inter-metallics) alone, has left myriad elemental blends unexplored.

(3) Limitations of computational resources – As pointed out above, the plausibility of any given stoichiometry is determined by the frequency with which that composition is visited by the imagination engine (i.e., the depth of the IE's attractor basins). Therefore, it is necessary to carry out thousands of experiments with any given composition ABC to establish the most likely distribution in stoichiometries x, y, and z. such runs would require months worth of CPU time on Pentium II class processors

and weeks on more powerful computational platforms such as Cray supercomputers. In this study, only ten possibilities for any given composition ABC were explored within each 3 day run of the Creativity Machine.

Nevertheless, the Creativity Machine Paradigm enjoys an adaptability that other discovery techniques totally lack, the ability to learn from its own mistakes. Therefore, we envision subsequent stages in this effort to test many of these materials predictions in the laboratory and to progressively improve on the accuracy of the Machine. Stoichiometries and properties manifesting significant error will be appended to the existing training exemplars and absorbed into any subsequent training of the component neural network modules.

4.2 Materials Discoveries

Potential Materials Discovery Having Technological Utility

In a preliminary exercise of the MatterHorn database, materials projected to possess exceptionally high or low values of the calculated properties are enumerated below. Of course the tabulations that follow do not exercise the full potential of the resulting materials discovery database to identify materials falling within intersecting properties regimes or families (i.e., substances having Tc above 20 deg K and Mohs Scale hardness above 9.5). Nevertheless, the following materials discoveries reveal compositions and stoichiometries that may exceed the known champions.

4.2.1 Refractory Materials

Ultra-Refractory Binaries – Compositions of the form RexHy, RexLiy, and LixWy show predicted melting points in excess of 4100 degrees C. The most refractory materials proposed by the CM include the entries shown in Table 3. Many of these materials, if shown to be viable, could find niche applications in the aerospace and nuclear fusion arenas. Materials projected to be especially stable, such as Re_5H_2, could be produced by a variety of techniques that include pulsed laser deposition, ion-implantation, or possible in situ implantation within nuclear radiation environments.

Table 3. Most Refractory Binary Systems

System	MP (deg K)	ΔG_f^0 (kcal/mole)
Re2H	4145	-41.3
Re3H	4144	-67.4
Re5H	4139	-110.3
Re6H	4136	-70.8
ReH2	4135	-26.6
Re2Li	4132	-41.3
Re3Li	4129	-66.0
Re5H2	4127	-369.3
Re5Li	4123	-105.5
ReLi2	4120	-26.3

Ultra-Refractory Ternaries – Some of the most refractory ternary systems discovered by the CM are listed in Table 4. All of these compositions appear to be thermodynamically stable. Synthesis routes for such ternary systems could include pulsed laser deposition, e-beam, or possible 'crucible chemistry' routes.

Table 4. Most Refractory Ternary Systems

System	MP (deg K)	ΔG_f^0 (kcal/mole)
B1K6W2	4160	-550.0
Al1Li7Ta2	4159	-550.0
Cm3Li2Mg	4151	-549.8
H6Li5Ta	4146	-550.0
Cm1Li4Mg	4139	-549.9
K4Li2Lu	4127	-550.0
H7LiTa	4125	-550.0
Cm2Li6Mg	4121	-549.9
CmLi5Mg2	4113	-549.8
Ba4He2Tm1	4094	-550.0

4.2.2 High Density Materials

High mass density films may be used to provide effective shielding for other systems within ionizing radiation environments. Satellite, medical, and weaponry systems could greatly benefit. Such materials could also be utilized in light weight armor schemes to protect against kinetic energy weaponry or to serve as replacements for the highly toxic inertial content of hyper-kinetic, armor-piercing projectiles.

One very likely route to the production of materials such as Pt_5Xe_2 (Table 5) would be ion-implantation of preexisting Pt thin films.

Table 5. Highest Density Binaries

System	DENSITY (Grams/cm^3)	ΔG_f^0 (kcal/mole)
Pd7Xe	36.1	-12.3
Pd6Xe6	36.1	-260
Pd6Xe7	36.1	-258.5
Pd2Xe5	35.5	-152.5
Fr2Pt	31.4	-195.3
Pt5Xe2	31.4	-294.8
Cd5H	30.6	-304.4
Fr5Pt3	30.6	-480.9
PaPt4	30.5	-307.2
Pd7Po2	29.3	+4.6

Synthesis routes to ultra-mass-dense ternaries could include e-beam, as well as pulsed laser deposition. Hydrogenation could be achieved via ion-implantation or growth within high background pressure H_2 environments.

Table 6. Highest Density Ternaries

System	DENSITY (Grams/cm^3)	ΔG_f^0 (kcal/mole)
Ba6H3Pt6	35.8	-550.0
Am2PtSi3	35.7	-549.2
CmPd2Sr2	35.7	-549.6
Cm7PdSr7	35.7	-549.9
Cm5Pd3Sr2	35.6	-549.8
PaPd4Tc7	35.5	-549.6
Cm5Pd5Sr2	35.4	-550.0
Ba5H5Pt2	35.4	-550.0
Fr7H3Pt	35.4	-550.0
Am5Pd3Sr3	35.4	-549.5

4.2.3 Ultra-Hard Materials

A casual perusal of Table 7 reveals the usual preponderance of both carbides and borides among the known ultra-hard materials. The interpretation of such materials is that of native cubic carbon or boron lattices dressed with interstitial impurities, typically of large atomic radius (Thaler, 1998). Some of these materials may be accessible via ion-implantation of heavy ions into either diamond or boron. In the case of diamond, implantation would be along the [110] direction of diamond. Furthermore, Laser Absorption Wave Deposition (LAWD, Thaler, 1993), could likewise be used on heavy metal laser ablation targets such as radium or thorium, or uranium to achieve the required ion-implantation energies. The resulting dressed diamond lattices could deliver the trans-diamond hardnesses through the stiffening of the native diamond lattice, provided graphitization is minimal.

Table 7. Ultra-Hard Binaries

System	HARDNESS (Mohs Scale)	ΔG_f^0 (kcal/mole)
C7Ra6	>9.9	-545.9
B5Th6	>9.9	-534.5
B7Pa7	>9.9	-541.7
B5Th7	>9.9	-533.1
B5Fr5	>9.9	-534.5
C6U5	>9.9	-544.0
Pa2Tl6	>9.9	-94.1
B6U7	>9.9	-540.6
B7U6	>9.9	-537.3
Th3Tl5	>9.9	-374.8

Synthesis routes for the potential ultra-hard ternaries listed below could proceed by e-beam, crucible chemistry, or pulsed laser deposition.

(Note that in considering the entries of Tables 7 and 8, the anticipated Mohs Scale accuracy is ± 1 Mohs Scale unit.)

Table 8. Ultra-Hard Ternaries

System	HARDNESS (Mohs Scale)	ΔG_f^0 (kcal/mole)
AlH4Yb	>9.9	-550.0
BHe7Yb4	>9.9	-550.0
C3Li7Lu2	>9.9	-550.0
C2Na3Yb2	>9.9	-550.0
BH6Hf7	>9.9	-550.0
He3Mn4S	>9.9	-549.3
C2Na4Yb4	>9.9	-550.0
Ge6LiYb7	>9.9	-549.9
CLi3Ta3	>9.9	-550.0
B6Ce5S3	>9.9	-549.4

4.2.4 Highly Refractive Materials

The ultra-refractive materials discoveries listed below, in Tables 9 and 10, could have potential utility in waveguide, optical fiber, and integrated optics applications. Both binary and ternary materials may be feasible through e-beam or pulsed laser deposition techniques.

Table 9. Highly Refractive Binaries

System	n_D	ΔG_f^0 (kcal/mole)
Au2Hg3	>4.2	-23.7
Au2Hg4	>4.2	-28.0
Au6Hg	>4.2	-85.1
AuGe2	>4.2	+7.6
AuHg5	>4.2	+33.1
Au4Hg4	>4.2	-371.8
OsPb	>4.2	-17.5
Ir2Sn	>4.2	-15.7
HgSn	>4.2	-4.5
Ir2Pb	>4.2	-1.7

Table 10. Highly Refractive Ternaries

System	n_D	ΔG_f^0 (kcal/mole)
CdPbSm	>4.2	-549.6
Mn2Sn2U4	>4.2	-549.7
RaTe5U6	>4.2	-549.7
Mn3Sb2U4	>4.2	-549.8
Mn3Sn2U2	>4.2	-549.8
In4PbPu4	>4.2	-549.9
Cr2NiU6	>4.2	-549.7
SnTcTh4	>4.2	-549.6
Nb5TeYb7	>4.2	-549.7
Ra7Te4U7	>4.2	-550.0

4.2.5 High-Temperature Superconductors

While not approaching the superconductive characteristics of the YBCOs, many binary and ternary compositions are promising relatively high critical temperatures, well above 20 degrees K (Tables 11 & 12). High energy argon-implantation of sodium, might be an inroad to the synthesis of such materials, which could be utilized in satellite and spacecraft applications.

Table 11. High Temperature Binary Superconductors (interpreted as interstitial solutions)

System	T_c (deg K)	ΔG_f^0 (kcal/mole)
Ar7H7	>21.3	-370.0
Ar7H6	>21.2	-448.1
Ar6H4	>21.0	-527.7
Ar7H2	>20.9	-486.1
Ar5Na6	>20.9	-413.7
Ar6Na3	>20.9	-537.6
Ar3Na5	>20.8	-391.5
Ar5H4	>20.8	-506.3
Ar6Na2	>20.8	-508.6
Ar6H2	>20.8	-419.2

Table 12. High Temperature Ternary Superconductors

System	T_c (deg K)	ΔG_f^0 (kcal/mole)
Cl7HeMg	>21.4	-546.9
F7HeMg3	>21.2	-544.9
SSeSi2	>21.0	-547.9
F6HeMg4	>21.0	-546.8
Ar7He3Mg3	>20.9	-548.5
B3BeC3	>20.9	-546.7
Ar6BMg2	>20.9	-546.1
NNaSi5	>20.9	-546.4
Ar5HeMg2	>20.9	-548.5
Cl6H6Li2	>20.9	-521.7

4.2.6 Highly Magnetic Susceptibility

The availability of extremely high magnetic susceptibility materials could have a profound effect upon a variety of industries that depend on high magnetic moment: (1) super-motors, (2) generators, (3) transformers, (4) radar-absorbing materials and structures, etc. Particularly exciting is the possibility of producing super-magnetic materials that circumvent the need for costly rare-earth constituents such as lanthanum (i.e., the high price of lanthanum in rare-earth iron boride supermagnets represents a prohibitive barrier in the economics of such devices). Therefore Fe_7Ne_7 and Cr_3Ne_5, (Table 13) potentially synthesized through ion-implantation of neon, into iron, or pulsed laser deposition, could offer single domain films that could be sandwiched into ultra-magnetic structures.

Table 13. High Magnetic Susceptibility Binary Systems

System	$X_{mag}^{(1E-6\ CGS)}$	ΔG_f^o (kcal/mole)
Fe7Ne7	578,000	-539.3
Ar3Yb	345,000	-455.4
Cl3Yb	344,000	-429.5
Fe7Ne6	333,000	-536.6
Ne4Yb	332,000	-488.0
Cr4Ne4	320,000	-521.4
Mn7Ne5	318,000	-530.2
Ar2Yb	311,000	-406.5
H7He6	308,000	-514.5
Cr3Ne5	303.000	-516.0

Table 14. High Magnetic Susceptibility Ternary Systems

System	$X_{mag}^{(1E-6\ CGS)}$	ΔG_f^o (kcal/mole)
C6Hf3I7	1,124,000	-549.2
C6Lu3Se5	1,063,000	-549.0
C7Lu3Te5	1,035,000	-549.5
Cl7Dy5Se5	1,005,000	-549
C7Er5I7	981,000	-549.4
Cl6Lu7Sn5	947,000	-549.5
C5Er6I6	925,000	-549.5
O7Sm7Te4	905,000	-549.5
H5Lu6Se6	903,000	-549.6
O6Sm7Te7	892,000	-549.5

4.2.7 High Electrical Conductivity Materials

If not superconductive, the following CM-theoretized systems (Tables 15 & 16) could offer negligible resistance to electrical current at room temperature. Since most of the constituent elements of the binary systems (e.g., Ac and Pt) are rare or costly, they could offer utility within microcircuitry. In contrast, it appears as though a number of ternary systems could offer the batch chemistry needed for large scale power transmission (i.e., Al_7Be_2C).

Table 15. High Electrical Conductivity Binary Systems

System	$\rho\ (\Omega\text{-m})$	ΔG_f^o (kcal/mole)
Bi2He	2.38E-9	+30.7
Ac5Pt5	2.39E-9	-483.6
Ac6Pt	2.40E-9	-350.6
Ac6Pt5	2.41E-9	-498.1
Ac3Pt4	2.41E-9	-472.8
Bi3He	2.42E-9	+18.0
Ac5Pt	2.43E-9	-306.8
AcPt7	2.44E-9	-68.9
La7Pt3	2.46E-9	-482.4
Ba7Pt3	2.50E-9	-497.2

Table 16. High Electrical Conductivity Ternary Systems

System	ρ(Ω-m)	ΔG$_f^0$ (kcal/mole)
Hf5Pd6Se2	8.24E-10	-550.0
B6Be6He3	8.30E-10	-520.8
Pu2Rb6Se4	8.94E-10	-549.3
Pu5Rb6Se6	9.20E-10	-549.5
Am4Na3Te2	9.30E-10	-549.9
B6Be5C	9.33E-10	-548.8
Al7Be2C	9.33E-10	-545.9
C5Ca6He2	9.42E-10	-504.9
B6He3Mg	9.47E-10	-549.6
Pu3Rb4Se	9.86E-10	-548.6

4.2.8 High Thermal Conductivity Systems

With the many frustrated attempts by industry to achieve diamond heat-sinking of microelectronics, the advent of alternative materials rivaling the thermal conductivity of that material (O(10E+3-10E+4 W/m-K)) could be a welcome development.. Below are listed several binary and ternary alternatives (Tables 17 & 18) drawn from the MatterHorn database. The presumed inert gas dressed boron and aluminum lattices could be achieved through ion-implantation, whereas much larger yield intermetallics such as Pt$_7$Rh$_4$ and the range of ternary systems recommended could be achieved through batch chemical routes.

Table 17. High Thermal Conductivity Binary Systems

System	σ$_T$ (mW/cm-K)	ΔG$_f^0$ (kcal/mole)
BNe	4.2E+4	-127.2
AlNe	4.1E+4	-127.8
BKr	4.1E+4	-37.5
IKr	3.9E+4	+37.1
AsBr	3.0E+4	+5.7
AsO	3.0E+4	-49.1
Cr4Mg	2.9E+4	-36.2
Pd7Re	2.9E+4	-159.6
Pd7Rh	2.8E+4	-379.0
Pt7Rh4	2.8E+4	-414.2

While the substituted inert gas systems, such as BNe appear difficult, if not costly to produce via implantation techniques, it appears that several batch chemistry candidates are likely and economical. These latter systems include AsBr, AsO, Cr$_4$Mg, or small stoichiometric variations upon these compositions.

Within the projected ternaries, Cu$_7$HTi , or a related composition, likewise promises to be a cost-effective batch chemistry candidate.

Table 18. High Thermal Conductivity Ternary Systems

System	σ_T (mW/cm-K)	ΔG_f^0 (kcal/mole)
Pt7RhSc	2.9E+4	-549.5
Pd7ScZr	2.9E+4	-549.7
Pd7PrTi	2.9E+4	-546.4
Pd7ScU	2.9E+4	-550.0
Pd5VYb	2.9E+4	-549.7
Pt7RbSi3	2.9E+4	-548.5
Pd7TaYb2	2.9E+4	-550.0
Cu7HTi	2.9E+4	-549.0
Pt6TiZr3	2.9E+4	-550.0
Pt6SiYb2	2.9E+4	-550.0

4.2.9 Low Coefficient of Thermal Expansion Systems

Highly stable optical resonantors and lasers could be fabricated from low CTE materials, such as those proposed below. With negligable CTE, many of the ternary systems Tables 19 &20) could form the basis of temperature resistant length standards.

Table 19. Low Coefficient of Thermal Expansion Binary Systems

System	CTE(1E-6/degK)	ΔG_f^0 (kcal/mole)
Au5Zr7	0.0	-39.3
Cl2Tc	0.0	-180.2
Li5Os7	0.0	-447.8
PrSe	0.0	-42.4
Pt7Pu5	0.0	-398.8
La2Ne2	-0.1	-212.2
ClSr	-0.1	-138.2
At2Br6	-0.1	+0.6
Cl3Tl2	-0.1	-235.4
Ar3Os2	-0.1	-367.9

Table 20. Low Coefficient of Thermal Expansion Ternary Systems

System	CTE(1E-6/degK)	ΔG_f^0 (kcal/mole)
Ar4 I3S5	0.0	-322.8
Br2La4Te	0.0	-542.8
Cu3Kr2S7	0.0	-452.9
As3Br2Te6	0.0	-510.3
NS2Sr4	0.0	-550.0
Al O2Sb5	0.0	-550.0
La4O1Te2	0.0	-549.8
N3SeTc2	0.0	-549.7
Ar5Ce4F	0.0	-540.0
He3Pd6Pt7	0.0	-550.0

4.3 Schema Analysis

One major advantage of having created such a large theoretical database of materials and accompanying properties is its breadth and completeness. In vivid contrast with more exacting, traditional databases that tabulate scattered materials and measurements, to great exactitude, the Matterhorn is complete and capable of revealing trends that would have been difficult to discern with the relatively sparse, empirical data. Therefore, it is possible to seek myriad relationships by experimenting with the various observables captured within the MatterHorn tabulation, in what might very likely be coined "virtual empiricism."

One example of this process may be seen in a previous materials discovery effort (Thaler, 1998) in which approximately 20,000 hypothetical binary systems were amassed, using the CM paradigm, along with anticipated Mohs Scale hardness, free energy of formation, and lattice constants. Using these estimated values, hardness was plotted versus free energy density, to reveal generalized trend lines that that obeyed the very case specific (i.e., tetrahedrally bonded semiconductors) Van Vechten Equation (Phillips, 1973). As predicted by that relationship, hardness and bulk modulus scaled as the lattice energy density and hence approximately as the inverse cube of inter-atomic spacings. In the case of the tetrahedrally bound semiconductors, just a few data points had been fit to the theoretical equation. However, in the latter instance, 4,000 neural network projected data points had been thusly plotted to reveal whole families of $1/r^3$ correlations with hardness, each determined by the covalent versus ionic nature of their bonding schemes. The lesson learned there was that even with the inevitable trade-off in accuracy, patterns are captured with sufficient precision that major trends, heretofore unobserved, may be gleaned from voluminous data one generation removed from direct measurement.

In a similar exercise of the newly generated ternary systems database, 32,000 hypothetical systems were used to create an auto-associative map among all of the projected materials properties tabulated within the MatterHorn database. Then, using a network skeletonization process introduced in the same 1998 paper, prune away the less significant connection weights from that net to reveal not only the most strongly correlated and perhaps causal factors, but to also expose key schema (i.e., how these factors conspire through hidden layer nodes to functionally determine each network output.

In Figures 10A-E that follow, the full auto-associative properties net was pruned, starting at the output node of interest, working back through the hidden layer, and then to the input layer, graphically deleting connection weights whose absolute value fell below 1.5. The surviving weights were color coded to show their relative significance to the overall mapping, with red representing values above 3, yellow, those between 2.5 and 3, green representing those between 2 and 2.5, light blue those between 2.5 and 2, and dark blue those values below 2.

The high level interpretation of each of these network skeletons may represent a technical paper on its own. However, I will quickly comment on just a couple examples to demonstrate how the revealed schema may connected to our current understanding of
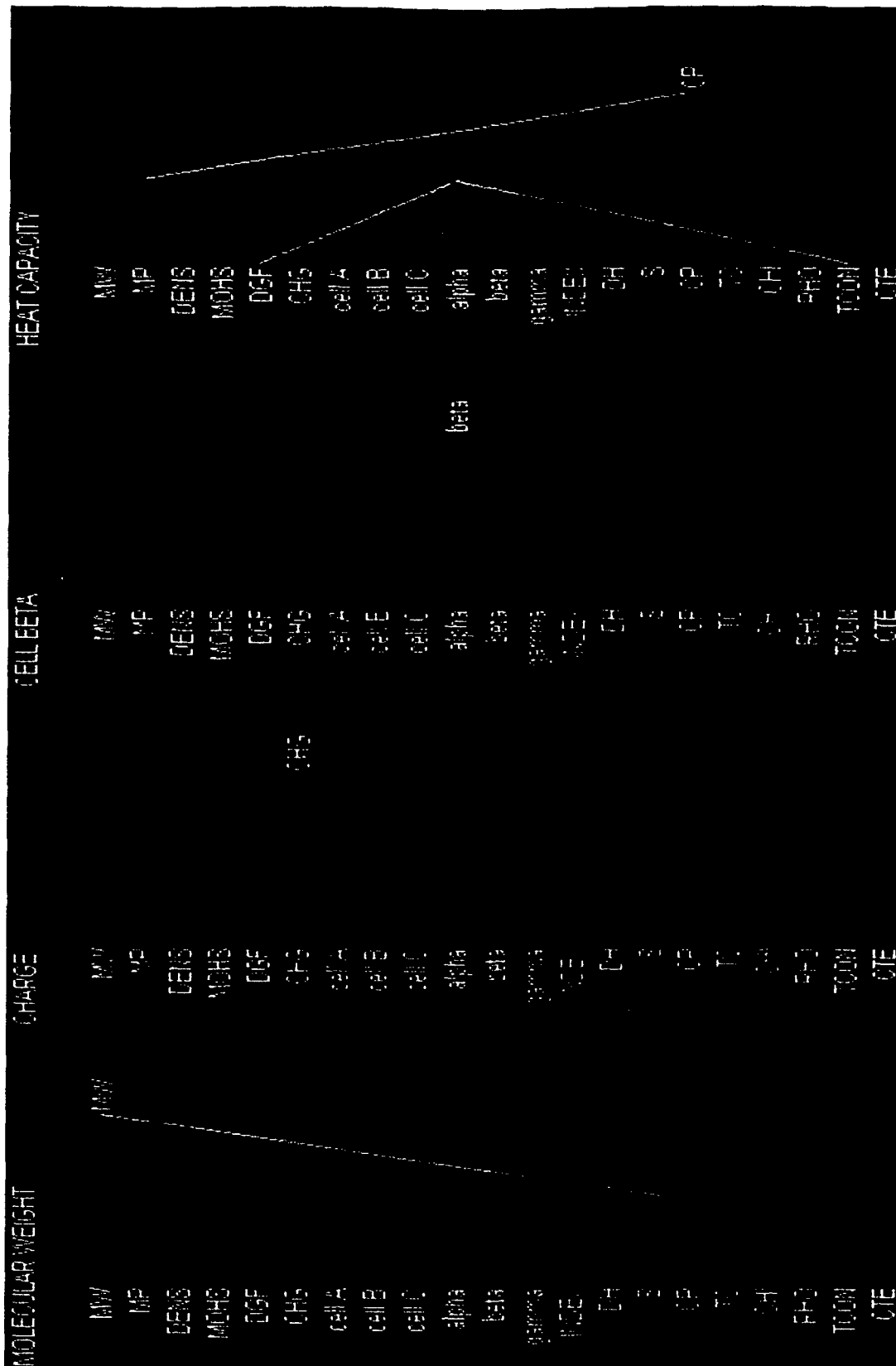
Figure 10A. Schema Analysis Via Network Skeletonization. Exposed connection traces yields factor analysis and the underlying schema connecting fundamental materials properties.
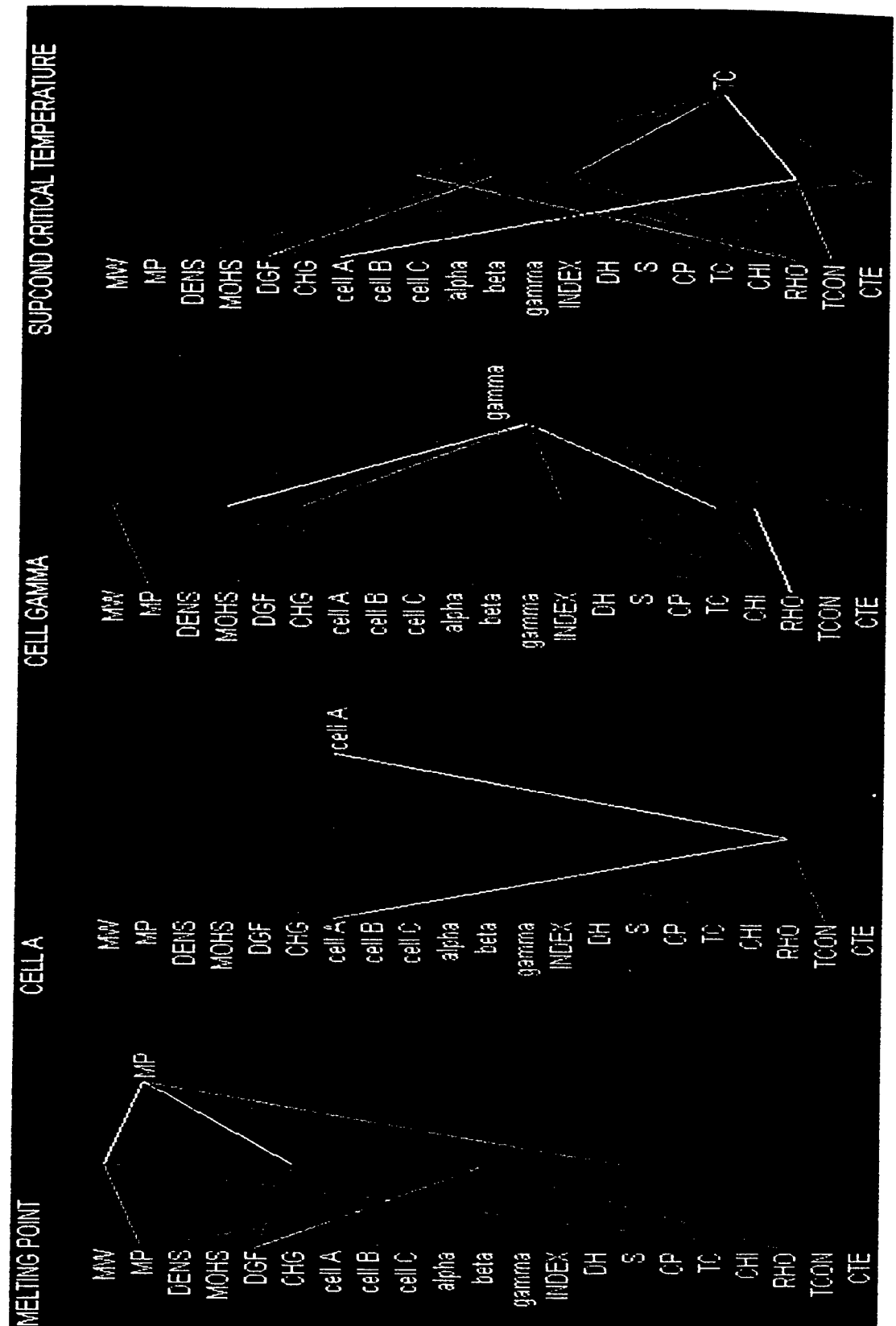
Figure 10B. Schema Analysis Via Network Skeletonization. Exposed connection traces yields factor analysis and the underlying schema connecting fundamental materials properties.
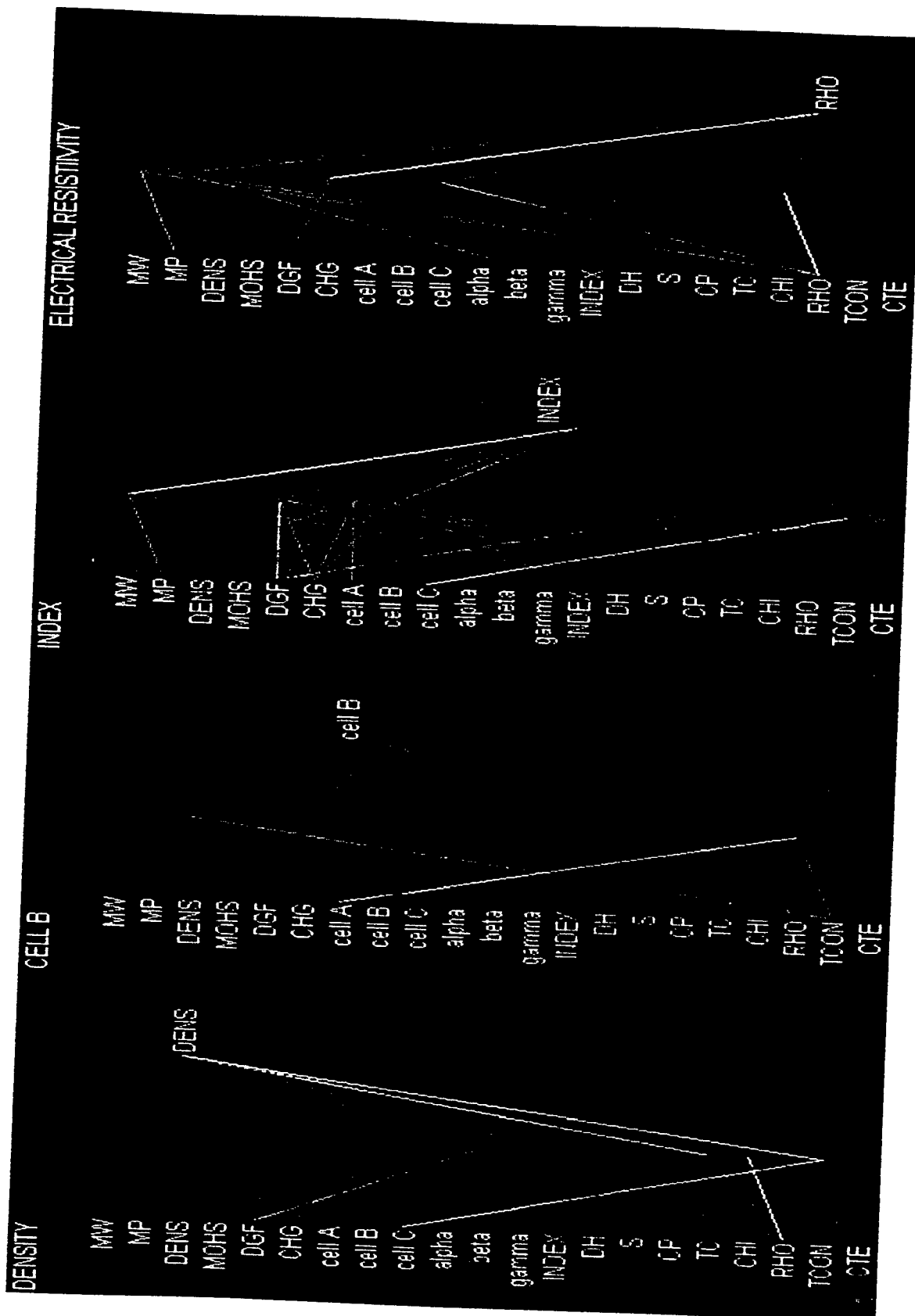
Figure 10C. Schema Analysis Via Network Skeletonization. Exposed connection traces yields factor analysis and the underlying schema connecting fundamental materials properties.
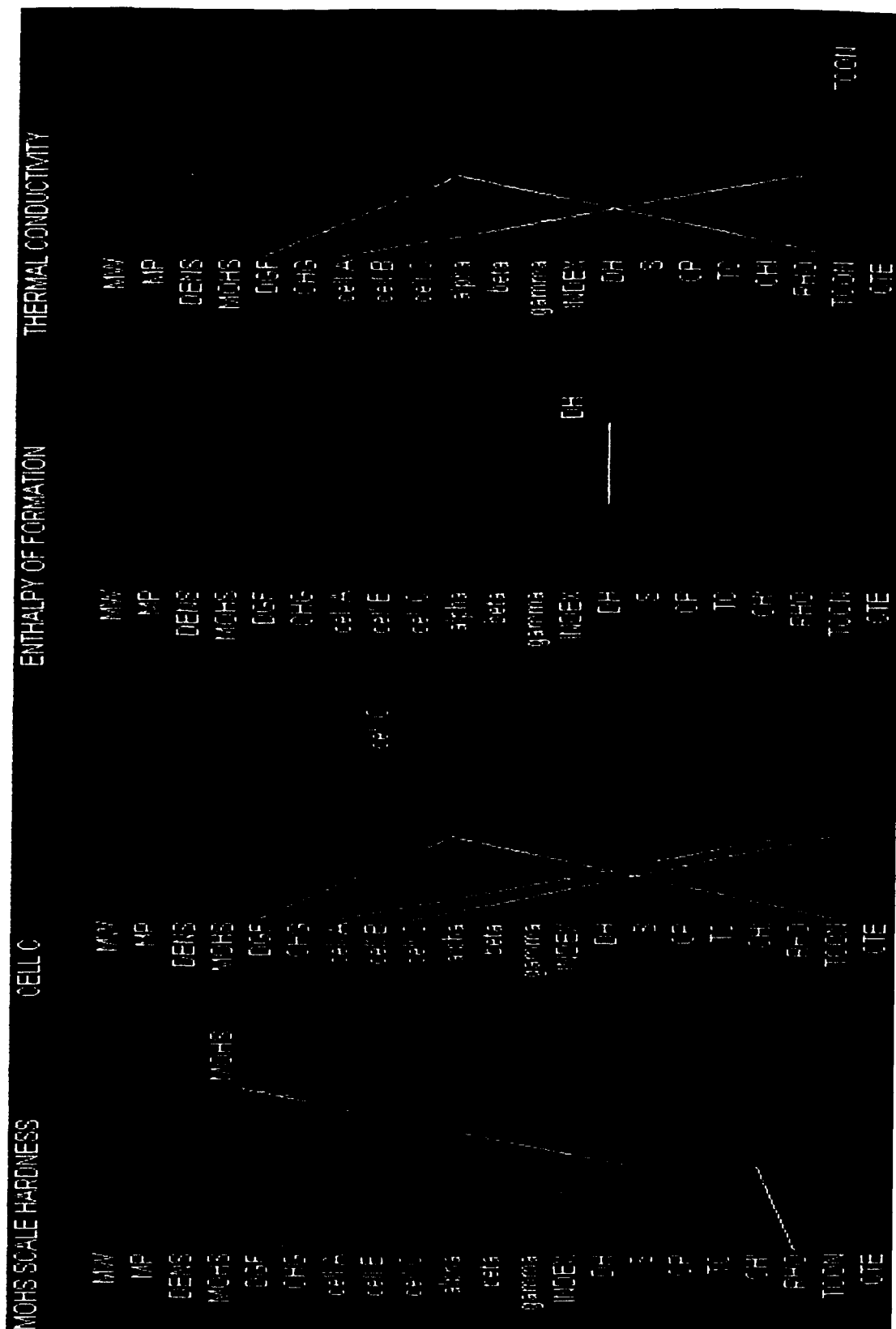
Figure 10D. Schema Analysis Via Network Skeletonization. Exposed connection traces yields factor analysis and the underlying schema connecting fundamental materials properties.

Figure 10E. Schema Analysis Via Network Skeletonization. Exposed connection traces yields factor analysis and the underlying schema connecting fundamental materials properties.
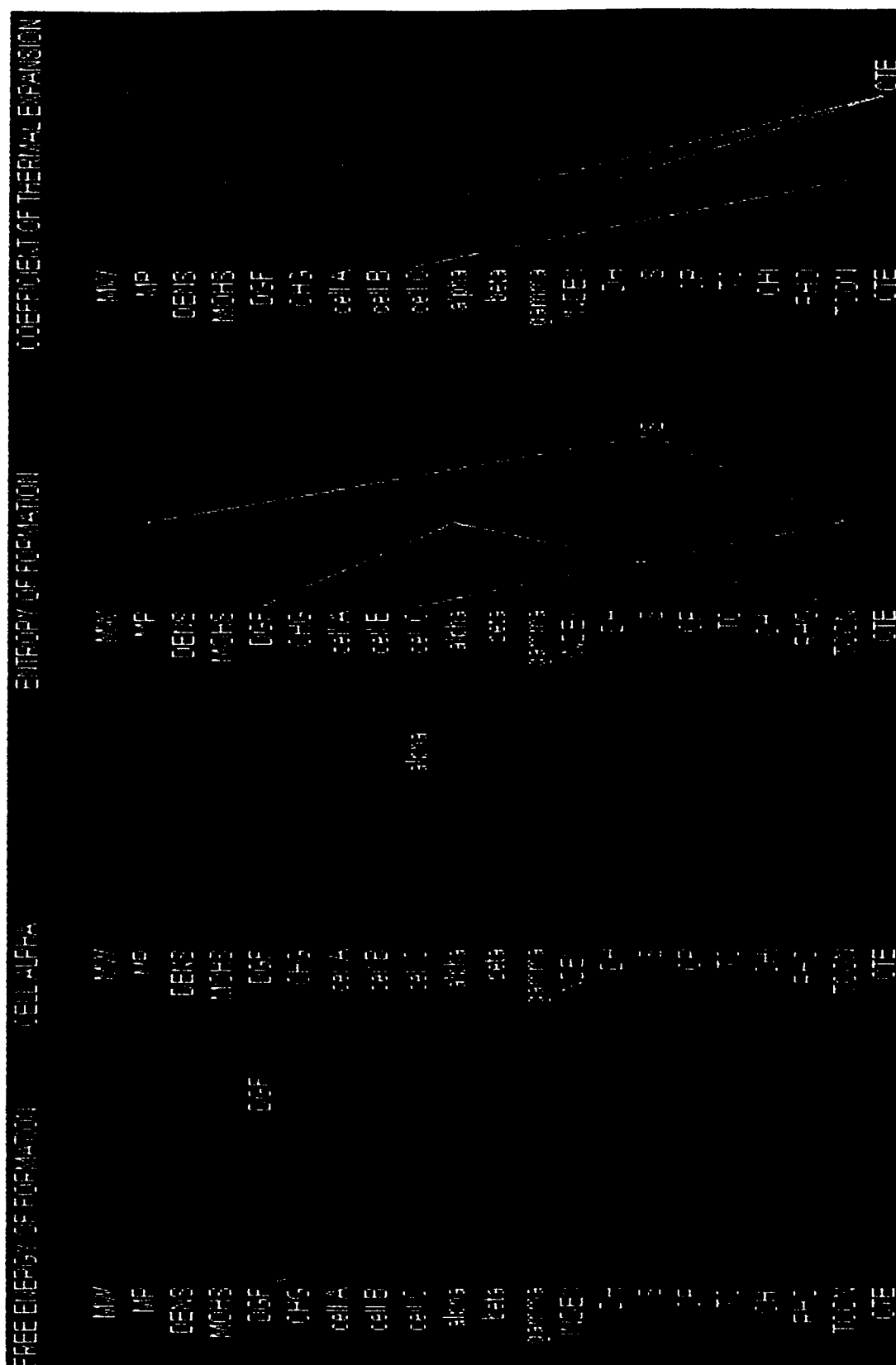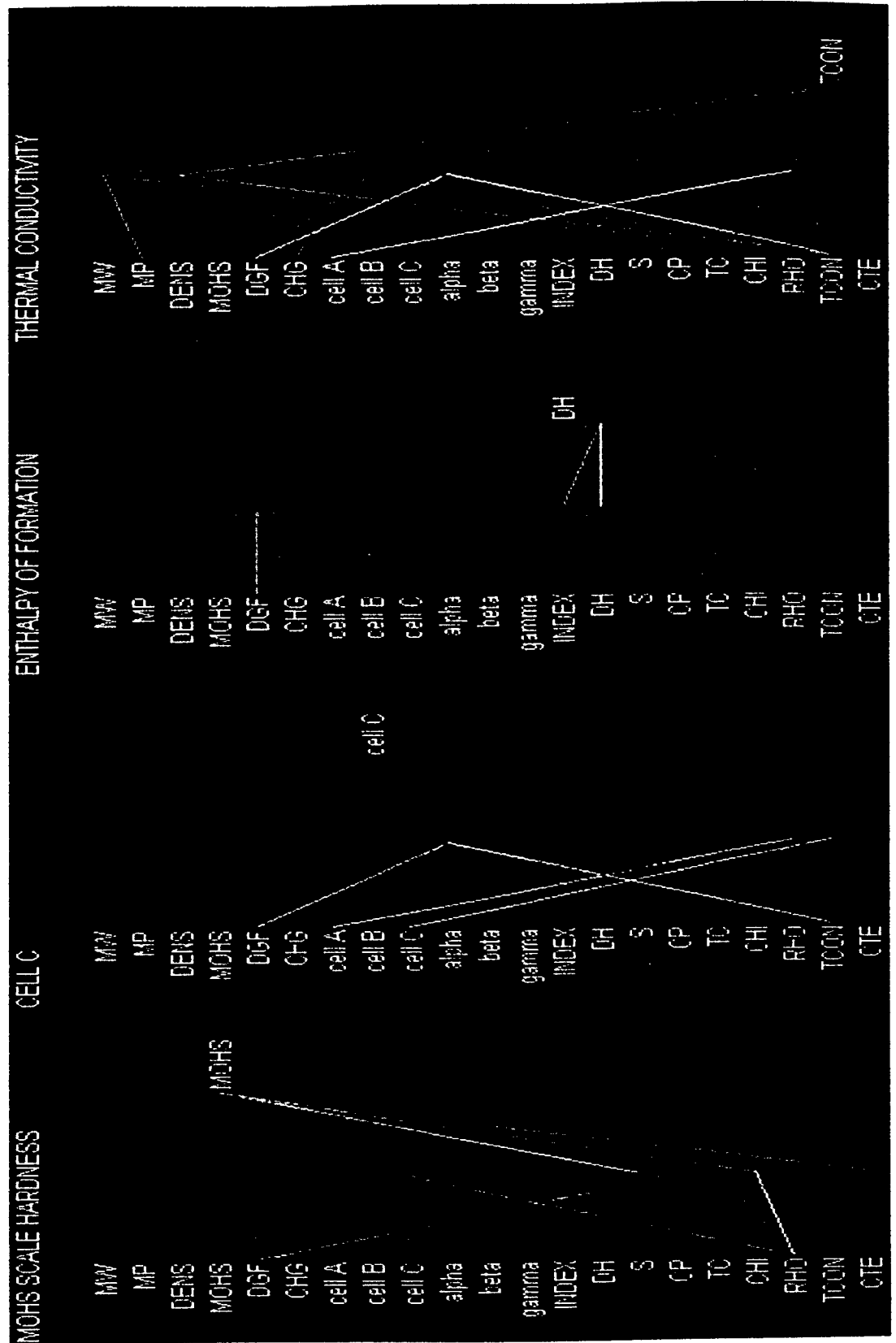
Figure 10D. Schema Analysis Via Network Skeletonization. Exposed connection traces yields factor analysis and the underlying schema connecting fundamental materials properties.

materials science.

Casual perusal of the skeletonization performed on the refractive index output (Figure 10C) reveals the major input factors to be free energy, lattice constants, electrical resistivity, magnetic susceptibility, and melting point. Viewing the refractive index as the sum over various atomic response functions, that primarily include interatomic potentials, we may rationalize the role of free energy (i.e., lattice energy) or melting point in functionally determining a primary component of the solid's oscillatory response function. We also note that free energy and lattice constants (that together determine interatomic spring constants) interact through an intermediate network node. Secondary factors include the response of electron plasma and magnetic dipoles through the resistivity and magnetic susceptibility, respectively.

Inspection of the thermal conductivity skeletonization shown in Figure 10D, implicates free energy and lattice constants being causal to this property. This schema would seem reasonable in that these properties conspire to determine interatomic potentials, as well as Brillouin zone characteristics. It is the flatness of the density of states, within this reciprocal space view, that in turn determines thermal conductivity.

The magnetic susceptibility skeletonization (not displayed) showed no significant connection traces between the input and output layers of the auto-associative map. This observation may corroborate the fact that highly magnetic materials consist typically of one paramagnetic atomic species, effectively supported within a diamagnetic host lattice. The latter host has little interaction with the former para- or ferromagnetic sub-lattice.

In short, we may repeat the analysis for each of the network skeletons of Figure 10, to not only corroborate much of what we know about the atomistic interpretation of materials behavior, but to fathom new and unforeseen relationships and causalities.

4.4 Desiderata

It goes without saying that the database produced by this materials discovery Creativity Machine would require several human lifetimes to fathom and corroborate. It will be criticized and ignored by many for its lack of precision, once contrasted against the tried and true empirical measurements of materials science. Furthermore, the 'computational accidents' comprising this database, will be largely overlooked, as flesh and blood scientists repeat and savor what has already been accomplished by machine.

In light of this overall discomfort with the notion of autonomous materials discovery, it may be constructive to think of the database produced through this Phase I effort as a slightly out-of-focus survey of all potential binary and ternary chemistry that human researchers may in turn mine for more exacting materials patterns: For example, successive queries of this discovery database may reveal the systems A_2B_3 and A_2B_5 as superlative niche performers, thus inspiring a series of rigorous laboratory investigations that single out A_2B_4 as the more plausible and/or superior material. In this sense, this expansive and speculative database is intended to drive materials tailoring and optimization rather than represent a chemical panacea and end-all.

Qualified by the above sentiments, allow me to point out that the Creativity Machine is a connectionist paradigm capable of learning from its own mistakes. Therefore, it is not a static software product, but an ongoing process that will, eventually bootstrap itself toward the precision of empirically gleaned materials databases. In this sense, the resulting improved database will progressively sharpen in focus until it rivals or exceeds the precision of traditional materials data resources. Those that heed its findings will acquire an extraordinarily powerful materials design tool.

5. References

Phillips, J. C. (1973). *Bonds and Bands in Semiconductors*, Academic Press, New York, NY.

Rowe, J. and Partridge, G. (1993). Creativity: A Survey of AI Approaches, *Artificial Intelligence Review*, 7, 43–70.

Rumelhart, D.E., Smolensky, P., McClelland, J.L., and Hinton, G.E. (1986). Schemata and Sequential Thought Processes in PDP Models, In *Parallel Distributed Processing, Exploration in the Microstructure of Cognition*, Volume 2: Psychological and Biological Models, MIT Press, Cambridge, MA, pp. 7–57.

Thaler, S.L. (1995). "Virtual Input Phenomena" Within the Death of a Simple Pattern Associator, *Neural Networks*, 8(1), 55–65.

Thaler, S.L. (1996a). Neural Nets That Create and Discover, *PC AI*, May/June, 16–21.

Thaler, S.L. (1996b). Is Neuronal Chaos the Source of Stream of Consciousness? In *Proceedings of the World Congress on Neural Networks*, (WCNN'96), Lawrence Erlbaum, Mawah, NJ.

Thaler, S.L. (1996c). A Proposed Symbolism for Network-Implemented Discovery Processes, In *Proceedings of the World Congress on Neural Networks*, (WCNN'96), Lawrence Erlbaum, Mawah, NJ.

Thaler, S.L. (1996d). Autonomous Materials Discovery Via Spreadsheet-Implemented Neural Network Cascades, *Journal of the Minerals, Metals, and Materials Society*, JOM-e, 49(4) [http://www.tms.org/pubs/journals/JOM/9704/Thaler]

Thaler, S.L. (1996e). Creativity via network cavitation – an architecture, implementation, and results, *Adaptive Distributive Parallel Computing Symposium*, Dayton, Ohio, 8-9 August, 1996.

Thaler, S. L. (1997). U.S. Patent 5,659,666, "Device for the autonomous generation of useful information", issued Aug. 17, 1997.

Thaler, S.L. (1998). Predicting ultra-hard binary compounds via cascaded auto- and hetero-associative neural newtorks, *Journal of Alloys and Compounds*, 279(1998), 47-59.

Thaler, S.L. and Conrad, D.M (1998). Real-Time Fault Detection Using Auto-associative Filtering, *AIRTC*, Oct. '98.

Thaler, S. L. (1998a). U.S. Patent 5,845,271. Non-Algorithmically implemented artificial neural networks and components thereof, issued Dec. 1, 1998

Thaler, S. L. (1998b). U.S. Patent 5,852,815, Neural network based database scanning system, issued Dec 22, 1998.

Thaler, S. L. (1998c). U.S. Patent 5,852,815, Neural network based prototyping system and method, issued Dec. 22, 1998.

6. Primary References Used For Training Exemplars

Cotton, F. A. and Wilkinson, G. (1972). *Advanced Inorganic Chemistry, 3rd edition*, John Wilet & Sons, Inc., New York, NY.

CRC Handbook of Chemistry and Physics, 53rd edition, 1972-1973, (ed. R. C. Weast), CRC Press, New York, NY.

CRC Handbook of Chemistry and Physics, 78th edition, 1997-1998, (ed. D. R. Lide), CRC Press, New York, NY.

Huheey, J. E. (1972), Inorganic Chemistry: Principles of Structure and Reactivity, Harper & Row, New York, NY.

Kiselyova, N. N. and Gladun, V. P. (1995). *Innovations in Materials Design, Parts I, II, & III: The Design of Inorganic Compounds*, Materials Process Design Branch, Materials Directorate, Wright Laboratory, Wright Patterson AFB, Ohio.

NIST Chemistry WebBook, National Institute of Standards and Technology, Gaithersburg, MD.

NIST/NRIM High Temperature Superconductors Database: Version 2.0, Standard Reference Data Program, National Institute of Standards and Technology, Gaithersburg, MD.

Pearson's Tables, (ed. P. Villars), ISBN: 0-87170-603-2, ASM Publication, 1997.
Pough, F. H (1960)., A Field Guide to Rocks and Minerals, 3rd edition, Houghton Mifflin Company, Boston, MA.

Shaffer, P. T (1996), Handbook of Advanced Ceramic Materials, Advanced Refractory Technologies, Inc.

TAPP Thermochemical and Physical Properties Database, ES Microwave.

Abrikosov, N. Kh., et al. (1969). *Semiconducting II-VI, IV-VI, and V-VI Compounds*, Plenum, New York, NY.

Summers, D. B. (1970). *Chemistry Handbook*, Willard Grant Press, Boston, MA.

Kittel, C. (1967). *Introduction to Solid State Physics, 3rd Edition*, John Wiley & Sons, New York, NY.

Ashcroft, N. W. and Mermin, N. D. (1976). *Solid State Physics*, Holt, Rinehart and Winston, New York, NY.

Gould, E. S. (1962). *Inorganic Reactions and Structure*, Revised edition, Holt, Rinehart and Winston, New York, NY.

McGraw-Hill Encyclopedia of Physics, 2nd Edition, (Ed. Sybil P. Parker), McGraw-Hill, New York, NY, 1993.